



PHONOLOGICAL CONTRAST AND PHONETIC VARIATION: THE CASE OF VELARS IN IWAIIDJA

JASON A. SHAW

Yale University

CHRISTOPHER CARIGNAN

University College London

TONYA G. AGOSTINI

University of Wollongong

ROBERT MAILHAMMER

Western Sydney University

MARK HARVEY

University of Newcastle

DONALD DERRICK

University of Canterbury

A field-based ultrasound and acoustic study of Iwaidja, an endangered Australian Aboriginal language, investigates the phonetic identity of nonnasal velar consonants in intervocalic position, where past work has proposed a [+continuant] vs. [–continuant] phonemic contrast. We analyze the putative contrast within a continuous phonetic space, defined by both acoustic and articulatory parameters, and find gradient variation: from more consonantal realizations, such as [u], to more vocalic realizations, such as [a]. The distribution of realizations across lexical items and speakers does not support the proposed phonemic contrast. This case illustrates how lenition that is both phonetically gradient and variable across speakers and words can give the illusion of a contextually restricted phonemic contrast.*

Keywords: Iwaidja, phonetic variation, lenition, ultrasound, manner, velar approximants

1. INTRODUCTION. One of the basic tasks in documenting a language is presenting an analysis of its segmental inventory. The ease of this task varies. Where there are minimal pairs and little phonetic overlap between segments, the task is straightforward, particularly when there is consistency across the speech community. However, even well-studied languages offer some challenging cases, such as the opposition between /n/ and /ŋ/ in English. The contrast is restricted to coda position, not the typical position for limited contrast (e.g. Becker et al. 2012, Beckman 1998), and it can be neutralized, for example, *walki*[n]/[ŋ], depending on the word, the phonological environment (e.g. /ten kɪnz/ → [tɛŋ kɪnz]), and the individual speaker (Fischer 1958). The distribution of [n] ~ [ŋ] is sociolinguistically complex and in ways that are not always recoverable in perception (Campbell-Kibler 2007), and it has been this way for at least hundreds of years.¹ While there are minimal pairs, such as *sin* /sɪn/ vs. *sing* /sɪŋ/, it is also conceivable that more limited corpora would not represent the contrast satisfactorily. We can

* We would like to thank the participants in this study, as well as Charlie Mangulda, David Cooper, James Cooper, Henry Guwiyul, †Ilijili Lamilami, Isobel Lamilami, and Maggi Maburnbi for sharing their languages and their insights on Iwaidja and Iwaidja culture with us. The dictionary that formed the basis of the work was originally assembled by Bruce Birch and Nick Evans. We thank them for developing this resource, for sharing it, and for many discussions about documenting Iwaidja. We also thank Bruce Birch, Mamaruni School, and West Arnhemshire Council for logistic support. Thanks also go to Laurence Bruggeman and Valeria Peretokina, who helped prepare visual stimuli for the study, and to the MARCS Institute for Brain, Behaviour and Development for technical support. The work benefited from audience feedback at Ultrafest VII, the 176th meeting of the Acoustical Society of America, the 15th Australian Languages Workshop, and the 96th meeting of the Linguistic Society of America. We would also like to thank Marianne Pouplier for comments on an earlier version of this article, as well as Jonah Katz, Andries Coetzee, Khalil Iskaraous, and three anonymous referees, whose comments greatly improved the final version. The research was supported by Australian Research Council Discovery Project 140100863, Language Documentation Grant 180717 from the Australian Research Council Centre of Excellence for the Dynamics of Language, European Research Council Advanced Grant 295573, and New Zealand Marsden Grant 12-UOC-081. Special thanks to the University of Canterbury's New Zealand Institute for Language, Brain & Behaviour for covering the open-source fees for this article.

¹ In his review of sociolinguistic work on (ING), Hazen (2008) cites Houston's (1985) discussion of an eighteenth-century spelling book mentioning different spellings for the following 'homophonous' pairs: *cof-fin* and *coughing*, *coming* and *cumin*, *heron* and *herring*.

thus imagine a documentation scenario in which a contrast goes undetected. In this article, we address the opposite scenario, in which undersampling phonetic variation across words and speakers leads to over-detection of contrast.

To make this scenario more concrete, consider a corpus of English containing one token of *thyme* produced as [tʰɪ̯m] and one token of *time* produced as [tʰɪ̯m]. In isolation (and taken at face value), such tokens provide minimal-pair evidence for a vowel length contrast. In this specific case, we know that the difference in vowel duration is likely attributable to lexical frequency (Gahl 2008), which influences word duration in homophone pairs (Lohmann 2018), or some factor correlated with frequency, such as age of acquisition (e.g. Carroll & White 1973) or local predictability (Cohen Priva & Jaeger 2018). In the general case, evidence for the conditioning factor may or may not be available in the corpus. For example, variation conditioned by sociolinguistic factors, such as speaker identity construction (Eckert 2008), may not be clearly recoverable from the variation in pronunciation alone (Campbell-Kibler 2009, Sherwood 2015). As a consequence, we may interpret the *thyme* [tʰɪ̯m] vs. *time* [tʰɪ̯m] opposition as (spurious) evidence for a vowel length contrast—until, that is, additional data is collected, and our analysis matures to the point that we can identify the alternative account.

For many languages where documentation is limited, phonological categorization across a range of phonetic realizations can present significant problems. Our specific case study involves how the phonetic space spanning from velar stop [k] to approximant [ɰ] to vowel [a] is partitioned into phonological categories in Iwaidja, an endangered Aboriginal language of Australia. This case presents an example of the general scenario sketched above. A range of phonetic variants at the velar place of articulation are attested. Our approach is to define a continuous phonetic space based on acoustic and articulatory dimensions relevant to variation in manner of articulation. Articulatory data collected in the field using ultrasound together with concurrent acoustic recordings contributed to the phonetic space. To inform our determination of categories within this continuous manner space, we use statistical methods for optimal clustering of phonetic data. We then investigate how the phonetic categories established based on bottom-up clustering of phonetic data map onto lexical items within and across speakers. The results indicate that while lexical items vary in how often they are realized with one phonetic category or another, the phonetic variants do not contrast in the standard sense. Rather, individual lexical items have tendencies toward one realization or another, and these tendencies vary across speakers. Although our concern is cast here as a problem for the linguist in the context of working on underdocumented languages, where undersampling is a concern, these same conditions may face a language learner with limited exposure, leading to language change across generations.

2. BACKGROUND.

2.1. IWAIIDJA. Iwaidja is an endangered Australian language whose traditional associations are in Northwestern Arnhem Land in Northern Australia. A generous estimate of the number of fluent speakers is around fifty. The four consultants for this article—CM, IL, ILM, JC—are first-language speakers of Iwaidja and long-term co-residents of Minjilang, the main community where Iwaidja is spoken as a significant daily language. All four consultants use Iwaidja as a main language for daily interaction. There are no reports of dialectal variation in Iwaidja.

The Iwaidja segmental inventory, as currently analyzed, is set out in Table 1 in IPA and in orthography.²

² Our default presentation of Iwaidja materials is with the standard orthography. We also use the standard Iwaidja orthography when presenting materials from other Australian languages, to facilitate comparison. We made this choice because the phonetic and phonological analyses of the velars are the points at issue.

Consonants					
	BILABIAL	ALVEOLAR	RETROFLEX	PALATAL	VELAR
STOP	p 	t <d>	ʈ <ɽ>	c <j>	k <k>
NASAL	m <m>	n <n>	ɳ <ɽ̃>	ɲ <ɲy>	ŋ <ŋg>
LATERAL		l <l>	ɭ <ɽ̣>		
FLAPPED LATERAL		ɭʼ <ɽ̣ʼ>	ɭʼ <ɽ̣ʼ>		
TAP		ɾ <ɽɾ>	ɽ <ɽɽ>		
APPROXIMANT	w <w>		ɻ <ɽ̣>	j <y>	ɰ <h>
Vowels					
		FRONT	CENTRAL	BACK	
HIGH		i <i>		u <u>	
LOW			a <a>		

TABLE 1. Iwaidja segmental inventory. Standard orthography given in <>.

As shown in Table 1, current analyses posit a phonemic contrast between a velar stop (orthographic <k>) and a velar approximant (orthographic <h>) (Evans 2009:160). The <h> category has also been described as a velar fricative (Pym & Larimore 1979:3–4). There are no velar fricative realizations in our data, but there are velar approximant realizations. Consequently, we follow Evans (2009) and categorize <h> as an approximant, /ɰ/ in IPA.

Critically, however, there is no one-to-one mapping between approximant realizations and the putative approximant phoneme <h>. The stop phoneme <k> is also commonly realized as an approximant. Figure 1 is a spectrogram of the word <kakurl> /kakuɭ/ [ɰaguɭ] ‘honey’ in utterance-medial position. The two instances of <k> in this recording of the word show different realizations. The word-initial <k>, which is in an intervocalic environment owing to the preceding /a/, is realized as a velar approximant [ɰ]; the word-medial <k>, also in an intervocalic environment, is realized more like a voiced stop [g].

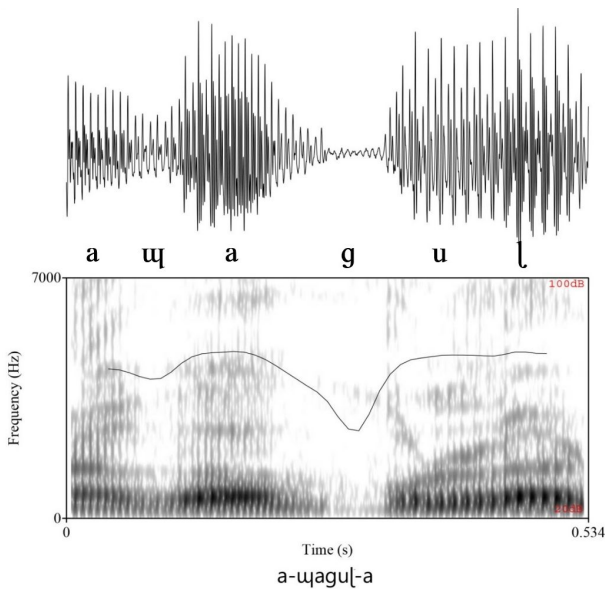


FIGURE 1. <kakurl> ‘honey’ in utterance-medial position with flanking low vowels. We have transcribed the first <k> as [ɰ] based on high intensity and spectral properties, and the second <k> as [g], due to the greater dip in intensity.

Measures of VOICE ONSET TIME (VOT) show that stops in all positions are generally voiced. Table 2 shows the average VOT values of 542 tokens collected from four native speakers of Iwaidja (Mailhammer et al. 2020).

	INITIAL	MEDIAL (V_V)	FINAL
VOT	−70 (50) ms	−66 (44) ms	−42 (44) ms

TABLE 2. Voice onset time (VOT) in Iwaidja in milliseconds (ms). Negative VOT values indicate that stops are voiced. Values are means, with standard deviation in parentheses.

The significant overlap in realizations between <h> and <k> accords with fieldworkers’ reports of persistent difficulties in distinguishing the two (our fieldwork; Birch, p.c.; Evans, p.c.). It is reflected in variant transcriptions of the same lexeme: for example, ‘collarbone’ is recorded as <kardalkardal> and <kardalh̥ardal>. By contrast, for the other stop vs. approximant oppositions— vs. <w>, <rt> vs. <r>, <j> vs. <y>—initial categorizations may involve some uncertainty, but further experience leads to persisting convergence in categorization.

The <h> segment has a restricted phonotactic distribution, appearing only morpheme-medially in the environment [+continuant] _ [+continuant]. As Iwaidja does not have onset consonant clusters, this means that <h> is found in two environments: (i) V]_{σσ}[_ V, and (ii) C_[+cont]]_{σσ}[_ V, where C is a coda consonant drawn from the set of <l, rl, rr>. The three other approximants, <w>, <r>, and <y>, appear both morpheme-initially and morpheme-medially.

Though <h> has a restricted distribution and there is a significant overlap in realizations with <k>, there are apparent minimal pairs contrasting <h> vs. <k>, as in 1.

- (1) a. r-aka
3SG.M>3SG.O-throw
‘He throws it.’

b. Ø-raha
3SG.S-be.upside.down
‘It is upside down.’

The persistent difficulty in both determining the phonetic variants and mapping them onto lexical contrast motivates the bottom-up phonetic approach to the data reported in this article and connects to some broader issues in the study of lenition, to which we now turn.

2.2. LENITION AND CONTRAST. Lenition comes in many forms—the term refers to both diachronic and synchronic processes that, from one perspective or another, involve ‘weakening’ (see Honeybone 2008 for an overview). The relevant notion of ‘weakening’ has been argued to be: phonological in nature, defined over representations that can be characterized in terms of ‘sonority’ (see Murray 1988 for an overview and Vennemann 1988 with inverse terminology labeled ‘consonantal strength’), ‘strength’ (e.g. Kaye & Harris 1990), or ‘effort’ (Kirchner 1998); phonetic in nature (e.g. Bauer 1988, Lavoie 2001); and/or diachronic in nature. The diachronic basis of ‘weakening’, which has been widely cited, was attributed to Theo Vennemann by Hyman (1975:165): ‘a segment X is said to be weaker than segment Y if Y goes through an X stage on its way to zero’. For example, on this definition, /ɣ/, the velar fricative, is weaker than /g/, the velar stop, since there exist sound changes that progress from /g/ → /ɣ/ → Ø, such as in Turkish (Sezer 1986, Ünal-Logacev et al. 2017); see also examples of velar lenition in §6.7.

There is debate about whether the range of phenomena characterized as lenition should receive a unified theoretical account. Unified accounts include that lenition involves reduction of segmental complexity (Kaye & Harris 1990), reduction of effort (Kirchner 1998), reduction of gesture magnitude and/or duration (Lavoie 2001), and failure to achieve a phonetic target (Bauer 2008). Other research has argued that a uni-

fied account of lenition is undesirable, as the processes that are grouped under the umbrella term *lenition* are heterogeneous in important ways. For example, Katz (2016) argues that a certain subset of lenition cases, which he terms *CONTINUITY LENITION*, are distinct in both their functional and phonological bases from other types of lenition, which he terms *LOSS LENITION*. Building on a proposal by Kingston (2008), Katz posits that continuity lenition serves a specific function related to the acoustic intensity profile of speech. On this proposal, the phonology actively changes segments so that the intensity profile of speech aligns with prosodic constituents. The proposed constraints require that acoustic intensity is low at prosodic margins relative to constituent-internal positions. Lower-intensity speech segments, such as plosives, are therefore preferred at prosodic margins, while higher-intensity speech segments, such as vowels, are preferred internal to prosodic constituents. Continuity lenition refers to cases of lenition that increase the acoustic intensity of the speech signal internal to prosodic constituents. For example, intervocalic spirantization ($/k/ \rightarrow [x] / V_V$) is classed as continuity lenition by Katz because the intensity of the interval $[VxV]$ is greater than that of $[VkV]$, owing to the difference in intensity between $[x]$ and $[k]$. Relative to $[k]$, $[x]$ has higher intensity and is therefore preferred when internal to a prosodic constituent. This contrasts with, for example, lenition of syllable-final $/k/$ in Indonesian (Cohen Priva 2017), which is loss lenition (not continuity lenition) because the process increases acoustic intensity at a domain margin.

Against the general backdrop of lenition characterized as weakening or laziness, the functional attribution of continuity lenition to improving the intensity profile of prosodic constituents stands out. On this view, continuity lenition is not weakening, because the function of reducing the constriction degree of consonants is to increase the intensity of the acoustic signal. Katz (2016) claims that continuity lenition has special phonological properties. The most important of these for our purposes is that continuity lenition does not neutralize contrast. Cases of lenition that do not meet the criteria for continuity lenition—that is, those that neutralize contrast and/or do not improve the continuity of high acoustic intensity within a prosodic word (or possibly other prosodic constituents)—are argued by Katz to behave like other (nonleniting) phonological processes. Therefore, if lenition is to have some special treatment in phonological theory, Katz suggests that such treatment need only apply to continuity lenition (cf. loss lenition).

From past work, it is not clear whether intervocalic velar lenition in Iwaidja is contrast-neutralizing. Certainly, $/k/$ can lenite to $[u]$, as shown in Fig. 1. Whether this lenition pattern is neutralizing hinges on whether $[u]$ has phonemic status. However, it is currently unclear whether $/u/$ really contrasts with $/k/$, as in the putative minimal pair in 1 above.

2.3. PHONETIC SPACES FOR MANNER OF ARTICULATION. At issue is how the range of phonetic variation at the velar place of articulation maps onto phonemic contrasts in Iwaidja. To address this question, we conducted a field-based ultrasound study. Articulatory data imaged using ultrasound was recorded along with concurrent acoustic data. From these articulatory and acoustic recordings, we constructed a two-dimensional phonetic space guided by both machine learning and theoretical concerns related to the case at hand. Essentially, the two phonetic dimensions that we employ in our analysis reflect the dimensions discussed in relation to allophonic variants of $/k/$ in Figure 2 below: change in intensity over time and constriction degree (i.e. stricture).

We used image-based statistical modeling to render sequences of ultrasound images as dynamic trajectories representing the degree to which tongue shape is differentiated from a vowel, $/a/$, which flanked the target consonants in all of our items. This tech-

nique effectively reduces the data to a single time-varying measure of vowel dissimilarity. Tracking change in vowel dissimilarity across target /...aCa.../ intervals reveals the degree to which the articulation of each consonant differs from that of the adjacent vowels.³ Extreme cases of reduction reveal no difference between the consonant and the adjacent vowels, while fully articulated consonants show sharp dips in the vowel dissimilarity signal.

Alongside our articulatory measure of vowel dissimilarity, we also computed the change in acoustic intensity across relevant /...aCa.../ intervals. Intensity provides an acoustic correlate of sonority and associated changes in manner (Parker 2008, 2017). A drop in acoustic intensity for the C in /...aCa.../ is expected; the degree of the intensity drop depends on the manner of the C. Since our data set includes both articulatory data and associated acoustic recordings, we are able to test the assumption of past work that intensity indexes constriction degree (e.g. Ennever et al. 2017). Ultimately, we show that the two-dimensional space we propose effectively captures continuous variation in manner from stops to vowels.

A schema of our proposed phonetic space is presented in Fig. 2. The vertical axis is defined by vowel (/a/) dissimilarity. This is based on the degree to which the consonant differs from /a/, as assessed by machine learning and classification of ultrasound images. Our approach represents the phonetic detail contained in sequences of tongue images as degree differences of vowel dissimilarity across target /...aCa.../ intervals. Larger dips in vowel similarity (greater dissimilarity) across this interval are indicated by larger numbers and correspond to articulatory configurations that differ to a greater degree from the flanking vowels (/a/ in this case). The second phonetic parameter is acoustic intensity. The horizontal axis of Fig. 2 is defined by the degree to which intensity drops across the same /...aCa.../ sequence over which vowel dissimilarity was calculated. Since both dimensions of the phonetic space are defined by dynamic parameters (degree of change over time), we can assess the status of intervocalic consonants without a need to pick out a specific moment in time or specific ultrasound image that corresponds to the putative target. This is important particularly in cases such as ours in which it is not clear what target we should be looking for or even, in extreme cases of reduction, whether there will be a target at all.

Within the two-dimensional manner space, we expect plosives to appear in the top right quadrant, as they are expected to differ articulatorily from flanking vowels to the greatest degree and also to have the largest dip in intensity, owing to tight constrictions in the vocal tract. The bottom left quadrant is where we expect to find vowels: these tokens are maximally similar in articulation to the flanking vowels and have the least dip in intensity across the /...aCa.../ sequences. The upper left quadrant and the lower right quadrant, as well as the center of the space, provide locations where we may observe approximants: tokens that have the articulatory configuration of vowels but nonetheless show a reduction in intensity (bottom right), or have flat intensity profiles but differ articulatorily from vowels (top left). If there is a perfect correlation between changes in constriction degree and changes in intensity, we expect to find our data falling in a di-

³ In order to capture possible differences between the medial /C/ and the flanking vowels, the /...aCa.../ interval was delineated by the nucleic-like steady-state locations within the first /a/ (beginning of the interval) and the second /a/ (end of the interval). These locations were first estimated automatically by locating the points of maximum intensity within the two halves of the /...aCa.../ segment; these time points were manually verified and hand-corrected as needed. In this way, the subsequent analyses allow for the possibility of both changes throughout the interval (i.e. consonant production) and no changes in the interval (i.e. full consonant lenition).

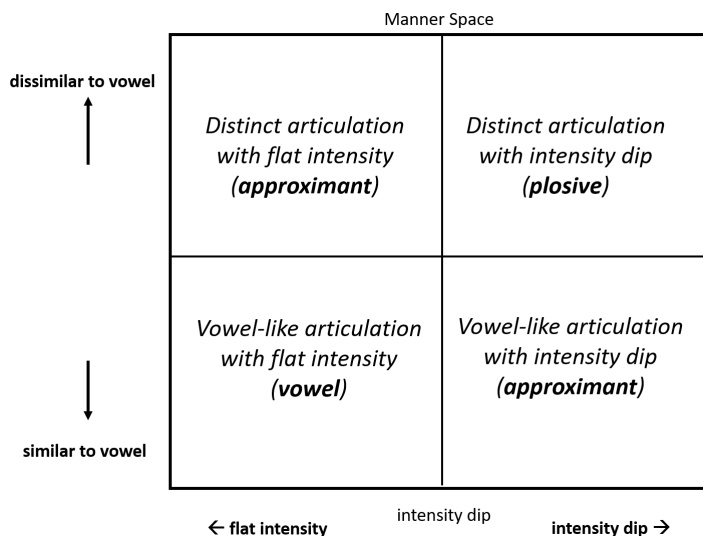


FIGURE 2. A two-parameter phonetic space in which manner of articulation ranges gradiently from plosives (top right) to approximants (top left and bottom right) to vowels (bottom left).

agonal from the lower left quadrant (vowels) to the upper right quadrant (plosives). In this case, we would find approximants in the center regions of the space.

After expressing our recordings of Iwaidja velar consonants within the phonetic space in Fig. 2, we proceed by examining the distribution of tokens within this space. We conducted a bottom-up classification of tokens based on the two phonetic dimensions, an approach that does not require that we know in advance the number of phonetic categories present in the data. As mentioned in our background section on Iwaidja above (§2.1), past analyses have posited a phonemic contrast between two velar consonants in the leniting environment.

3. EXPERIMENTAL METHODS. To assess the status of the proposed <h> vs. <k> opposition and the behavior of velars in leniting environments, we conducted a field-based experiment, collecting ultrasound and acoustic data from four native speakers of Iwaidja.

3.1. MATERIALS. All stimulus items were drawn from the current version of the Iwaidja dictionary. For practical reasons associated with elicitation ease, we focused on picturable nouns: ‘flying fox’, ‘sugar glider’, ‘saucepan’, and so forth. The full list of items for this study is provided in Table 3. Primary stress is on the first syllable in all of these items, which is the standard metrical pattern for unprefixated nominals in Iwaidja. These include eight picturable nouns with <h> flanked by /a/ on both sides (first column) and eight similar nouns with <k> in the same context (second column from the left). These are the items of primary interest. In addition, as a control comparison, we also selected items with <j> and <y> in the same intervocalic /...a _ a.../ context. We selected this pair of consonants because there is no question that <j> and <y> ([c] and [j]) are contrastive in this (and other) environments. Alongside our target <h> ~ <k> and <y> ~ <j> pairs, we also list (rightmost column) words that contain <a> in the first syllable (stressed position).⁴ The stressed <a> words were used to provide a baseline for

⁴ Some of the <a> words overlap with words listed elsewhere in the table. For example, <bakay> is listed in the second column because it contains <k> in the /a _ a/ context and in the last column because it contains <a> in stressed position

investigating vowel duration (see §5.4). We chose only stressed /a/ for this purpose in order to be conservative, since it is possible that Iwaidja is like many other languages in that stressed /a/ is longer in duration and has a more extreme articulatory target than unstressed /a/. The items in Table 3 were recorded along with fillers and words included for other studies, a total of seventy-five words in all.

<aha>	<aka>	<aja>	<aya>	<a>
<i>baharl</i>	<i>jalakaraj</i>	<i>bamaja</i>	<i>ardayangkayurl</i>	<i>bakay</i>
'3SG head'	'fish spear'	'saucepan'	'file snake'	'spike'
<i>kalahala</i>	<i>kalakalak</i>	<i>birraja</i>	<i>balayarrang</i>	<i>balayarrang</i>
'catfish'	'bird (generic)'	'rice'	'mast'	'mast'
<i>nahalak</i>	<i>dakamul</i>	<i>imajak</i>	<i>mayak</i>	<i>banikin</i>
'flying fox'	'throwing stick'	'wing'	'vine species'	'container'
<i>nawahaj</i>	<i>warraka</i>	<i>kajakajak</i>	<i>kayarlkayarl</i>	<i>dadbihi</i>
'subsection term'	'water lily'	'frog (generic)'	'firefly'	'river'
<i>wahardi</i>	<i>wakarti</i>	<i>kajalat</i>	<i>marriyalayala</i>	<i>dakamul</i>
'elbow'	'fighting club'	'mullet'	'freshwater crocodile'	'throwing stick'
<i>wahay</i>	<i>bakay</i>	<i>kalajarr</i>	<i>nayangkany</i>	<i>kajakajak</i>
'sugar glider'	'spike'	'ear'	'spear thrower'	'frog (generic)'
		<i>najabin</i>	<i>ildahaya</i>	<i>kalajarr</i>
		'barracuda'	'blue salmon'	'ear'
		<i>wubaja</i>	<i>walaya</i>	<i>kalahala</i>
		'police officer'	'cliff'	'catfish species'
			<i>waya</i>	<i>kaluku</i>
			'wire'	'coconut'
			<i>warrayaj</i>	<i>kandijawa</i>
			'red-winged parrot'	'flour, damper, bread'

TABLE 3. Stimulus items analyzed for this study in phoneme-based blocks.

For each stimulus item we sourced a representative photograph of the noun and assembled a visual presentation in Microsoft Office PowerPoint (2010), which displayed each photographic image combined with an orthographic representation in the same set order for each participant.

3.2. PARTICIPANTS. Four native speakers of Iwaidja participated in the study. At the time of the recording (2015), the speakers were between forty-two and eighty years old (one male aged forty-three years (JC), one female aged forty-two years (IL), one male (CM) eighty years, one female (ILM) seventy-nine years).⁵ This corresponds to 8% of the estimated total population of proficient Iwaidja speakers. All speakers live in Minjilang on Croker Island; they are all multilingual and use Iwaidja daily as one of their main languages. Except for the younger female, all participants are functionally illiterate. All participants reported having normal vision and no significant impairments that would impact articulation, hearing, or cognition.

3.3. EQUIPMENT. We collected ultrasound images using a GE LOGIQ E ultrasound, 2011, with an 8C RS wide-band microconvex array transducer. The probe head is 22 mm long and round, allowing for 180 degrees field of view. Relative machine settings were: frequency: 10 MHz; angle of capture: 118 degrees; depth CM; capture rate: 21 Hz. A VGA2USB Pro frame grabber was used to connect the ultrasound machine's VGA output to the data collection computers. Simultaneously, a Countryman omni-

⁵ ILM is the biological mother of IL.

directional lapel microphone was connected to a Sennheiser wireless receiver via a USBPre 2 preamplifier. FFMPEG software was used to capture ultrasound video and audio from the microphone.

The ultrasound probe was stabilized against the participant's jaw using a nonmetallic probe holder prototype described in Derrick et al. 2015, as well as an improvised probe holder with a bicycle helmet and a RØDE shotgun microphone shock mount through which the probe was fixated (for JC and IL). Audio and ultrasound synchronization were accomplished in post-processing using the manual method described in Carignan 2018 (see also Heyne et al. 2016).

3.4. PROCEDURE. Data was collected in different places on Croker Island and at different times in 2015. Two speakers, ILM and CM, were recorded at Adjamarduku outstation in April 2015. For these two speakers, the Derrick et al. (2015) probe holder was available. The other two speakers were recorded at Mamaruni School in November 2015. These speakers were recorded with the improvised probe holder, which proved reliable in stabilizing the probe. All recordings have a signal-to-noise ratio (SNR) of 35 decibels (dB), which is high quality for field recordings.

Participants sat in front of a computer screen. The experiment was preceded by a familiarization period during which the stimulus material was checked with each participant. Once the stimulus items were confirmed by the participants, they were familiarized with the carrier phrase *ngadbin _____ mana ba Iwaidja* 'we (excl.) say _____ in Iwaidja'. After familiarization, the ultrasound probe was affixed midsagittally; see Figure 3. The experiment began and concluded with a sequence of five voiceless alveolar stops combined with a low vowel (i.e. /ta ta ta ta ta/) to provide salient landmarks in the acoustic and articulatory signals to assist with later synchronization of the audio and ultrasound image data.

Words represented in the dictionary with <h> were recorded in a single block so as to encourage consistency in pronunciation. Words represented in the dictionary with <k> were also recorded together, in a single block. The same blocking strategy was used for <y> words and for <j> words as well. Tokens with /a/ were interspersed throughout all of the blocks. Participants produced each word in the carrier phrase at least two times within a block. Some speakers were willing to produce the words more than two times. The number of repetitions per target word per speaker is given in Table 4.

4. DATA ANALYSIS.

4.1. LIMITATIONS OF 'TARGET' FRAME ULTRASOUND ANALYSIS FOR OUR DATA. Figure 4 shows an example of the waveform, spectrogram, and ultrasound frames for one token. The large rectangle indicates the /aCa/ interval, which was the target of our analyses. The ultrasound waves reflect the changes in tissue density between the surface of the tongue and the air above the surface of the tongue.

A fairly standard approach to analyzing ultrasound images of the tongue (Fig. 4) is to trace the tongue edges from individual frames of ultrasound video (Li et al. 2005) and to analyze the differences between splines fit to the tongue-surface tracing in particular frames chosen to represent the achievement of an articulatory target (Davidson 2006) or to analyze some particular portion of the spline, such as the highest point, within a target frame (Noiray et al. 2014) or the distance from the probe to the tongue back surface (Recasens & Rodríguez 2016). The height of the highest portion of the tongue in the target frame would be, in principle, a useful measure for our purposes. To the extent that /k/ and /u/ contrast with /a/, we would expect that the tongue would be highest for /k/, followed by /u/ and then /a/. However, there are two issues that make this approach



a. Nonmetallic custom-made probe holder (Derrick et al. 2015).
b. Improvised nonmetallic probe holder.

FIGURE 3. Probe holders. (a) Custom-made probe holder, used for two subjects. (b) Improvised probe holder, used for two different subjects. (All photos used with permission.)

SPEAKER	REPETITIONS PER TARGET WORD	TOTAL # OF TARGET TOKENS
CM	2	24
IL	4–5	46
ILM	4–8	51
JC	3–6	49
TOTAL		170

TABLE 4. Stimulus items analyzed for this study by participant.

unsuitable for this study. Before describing the details of our approach, we first elaborate on why the standard approach of selecting target frames for analysis and measuring aspects of those target frames is suboptimal for this study.

The first issue is that we do not have any way to represent traces of the tongue edge relative to the hard structures of the vocal tract. In laboratory settings, this is sometimes done by tracking points on the head and on the ultrasound probe. After the data is collected, the tracings of tongue edges are rotated so that each frame is in a consistent orientation relative to the markers on the head; this corrects for any movement of the probe relative to the head across trials over the course of the experiment (Whalen et al. 2005). In the absence of this correction, small differences in the orientation of the probe relative to the jaw can introduce variability in the orientation of the tongue tracing, which

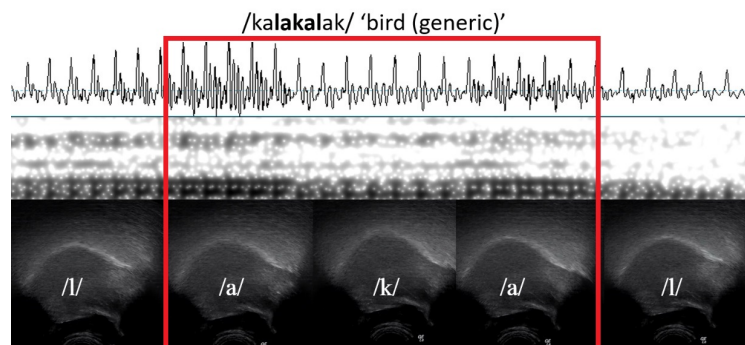


FIGURE 4. Acoustic waveform (top panel), spectrogram (middle panel), and ultrasound frames (bottom panel) for the /lakal/ portion of a token of /kalakalak/ 'bird'. The box indicates the /aCa/ portion of the signal that was the basis of our dynamic analyses of the acoustic and articulatory signals.

impacts the location of the highest-position measurement (Noiray et al. 2014). In the field setting, we were unable to track the relation of the probe relative to the head. This makes interpreting the tongue edge in absolute terms, such as highest point, somewhat unreliable. We cannot be certain that the highest point of the tongue within the ultrasound frame is really the highest point of the tongue in the vocal tract because the highest point depends on the exact orientation of the probe relative to the head.

A second issue is that—since we are investigating consonant articulation in a lenition environment—it was sometimes difficult to pick out a target frame. In some of the /...aCa.../ intervals, the tongue shape, indicated by tracings of the ultrasound frame, stayed very close to /a/ throughout, making it difficult to identify a frame that represents the consonant target (by raising of the tongue toward a velar target). We ultimately develop a new method to analyze the data that addresses the two issues above. Before moving on to the new method, we describe some tongue traces of 'target' frames using the standard method. The ultrasound frames with the highest tongue-body position within the /...aCa.../ interval were selected as the consonant 'target' for each token. Figure 5 shows raw tracings of the tongue edge, extracted using Edgetrack (Li et al. 2005), for one speaker (CM). The top panel shows markers and colors according to the presumed phonological category for each token—orthographic <k> is labeled as 'k'; orthographic <h> is labeled as 'u'.⁶ These are overlaid with tokens of /a/. The bottom left panel shows the highest point of the tongue in two-dimensional space. The bottom right panel plots Gaussian probability distribution functions fit to the vertical and horizontal position of the highest point of the tongue tracing. The axes of the plots are labeled as 'horizontal' position and 'vertical' position. Of course, since we were not able to correct for any possible variation in probe position across trials, these axes are 'vertical' and 'horizontal' only with respect to the ultrasound image and not necessarily with respect to vocal-tract anatomy. The units are in millimeters (mm), but the origin of the plots is arbitrary for the purposes of interpreting the data.

With the above caveats, we can see that for this speaker (CM) the range of variation for /a/ largely encompasses that of both /k/ and /u/. On average, /k/ is slightly higher in the ultrasound frame than /u/ or /a/, and /u/ is slightly backer in the frame than /k/ or /a/. Unfortunately, the dimensions that we would like to interpret, such as tongue

⁶ Note that this figure and several of the others are presented in full color in the electronic versions of this article, but in black and white in the print version; color versions of the figures are also available open access at <http://muse.jhu.edu/resolve/104>. Where colors are referenced in the print article, blue corresponds to dark gray, red to medium gray, and green to light gray.

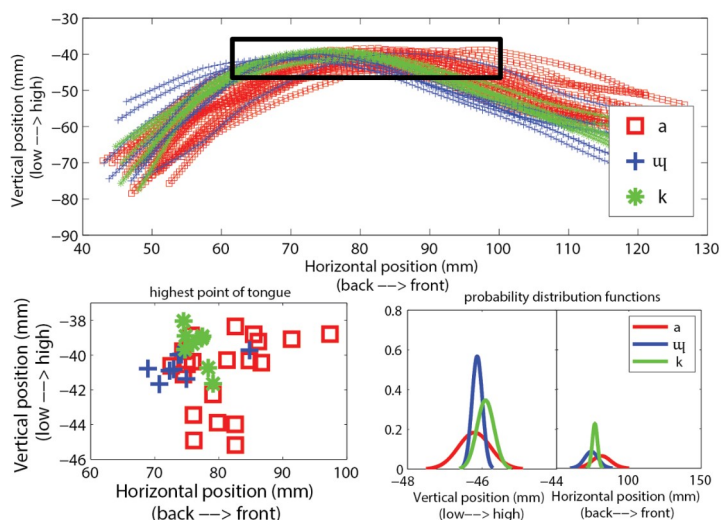


FIGURE 5. Velars compared to /a/: tongue tracings from ultrasound images overlaid (top panel); the highest point in the frame of each tongue tracing (bottom, left); probability distribution functions fit to the highest point of the tongue (bottom, right).

height, are precisely those that are somewhat unreliable in this data. The /k/ tokens were elicited in a separate block from the /u/ tokens (see §3.4), while the /a/ tokens were elicited across blocks. While it is reasonable to interpret the data at face value—that is, /a/ is more variable than velar consonants—an alternative possibility is that the angle of the probe tilted slightly between the blocks with /u/ (orthographic <h>) and the blocks with /k/. Rotating all of the /u/ tongue tracings slightly forward would change the location in the tracing that serves as the highest point in the frame, reducing the (already small) differences observed in backness and height between /k/ and /u/.

Part of our design was to include palatal consonants, in part because comparing the palatal stop and palatal glide would serve as an appropriate check on the methods. Figure 6 shows Edgetrack tracings of the palatal consonants along with the vowel /i/ from the same speaker (CM) as Fig. 5. In contrast to the velar consonants, there is more consistency for both palatal consonants—note that the horizontal axis spans 20 mm for the palatals (cf. 40 mm for the velars). The highest point of the palatal stops tends to be more anterior than that of the glides, in the direction of contact with the upper portion of the alveolar ridge. As with the velars, the palatal glides are more posterior than the palatal stops. Overall, the palatal data patterns largely as expected and is less variable than the velar data. This strongly suggests that at least part of the variability observed for the velars is due to real variability in articulation as opposed to noise in the field-based data acquisition process. It would be quite a coincidence (and maximally unlucky) for the probe orientation to shift just when the stimuli presentation changed from the <k> block to the <h> block and again for the <j> block and again for the <y> block. On the whole, we do not have any particular reason to suspect that we had substantial probe movement in any of our sessions. However, since we also do not have any method to rule this possibility out, we have developed an analysis that does not rely on interpreting differences across tokens (and blocks), which we hope will prove useful in analyzing both laboratory- and field-based ultrasound acquisition beyond the current study.

We pursue an analytical approach to the ultrasound data that minimizes possible confounds of probe position across speakers. The approach takes the whole temporal inter-

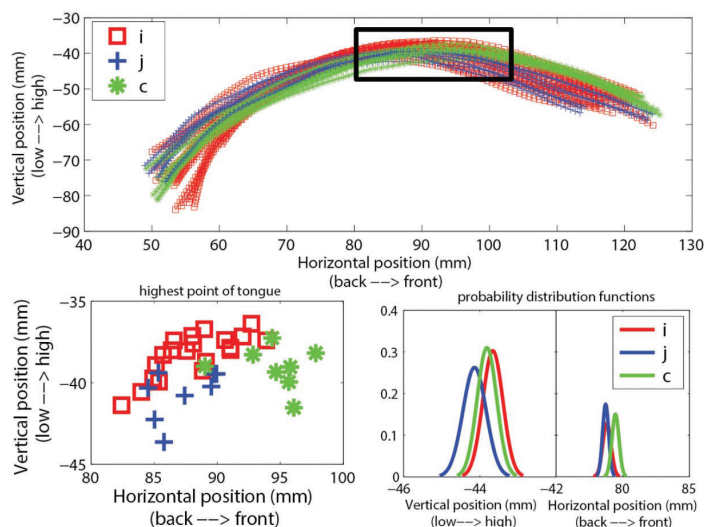


FIGURE 6. Palatals compared to /i/: tongue tracings from ultrasound images overlaid (top panel); the highest point in the frame of each tongue tracing (bottom, left); probability distribution functions fit to the highest point of the tongue (bottom, right).

val spanning /aCa/ into account, which has the added bonus that it is not necessary to select specific frames as targets. Our analysis compares the relation between frames within the temporal interval. The basis for analyzing frames is a LINEAR DISCRIMINATE ANALYSIS (LDA). The LDA classifier was trained on all target frames of /a/ produced by each speaker (more details below). Each frame is analyzed relative to adjacent frames. Ultimately, it is the relationship between frames within a token that serves as our main dependent variable. Analyzing this relationship (instead of picking out frames for analysis across tokens and speakers) has the advantage that the probe position need only be consistent within a single token. We also find that because we compute relative changes within-token against a speaker-specific model, the method effectively normalizes for speaker differences. As an additional bonus, the relational measure of the ultrasound data that we compute has properties similar to the relational measure of acoustic intensity that we use to codetermine phonetic category membership.

4.2. PHONETIC PARAMETERS. Our analysis includes articulatory measurements from ultrasound recordings as well as acoustic measurements from concurrently recorded audio (acoustic intensity). In both cases, we analyze changes in the signal over time across an /aCa/ interval, where C draws from either the palatal consonants, <y> and <j>, or the velar consonants, <h> and <k>.

DIP IN ACOUSTIC INTENSITY. The primary acoustic parameter of concern was the intensity contour across the target interval. Acoustic intensity was computed as amplitude (dB) in Praat (Boersma & Weenink 2013). Measurements were extracted from 6 ms intervals across the target /...aCa.../ sequences. A second-order polynomial was then fitted to the dynamically changing intensity signal within a token. The quadratic term of the equation captures variation in intensity across the /aCa/ interval. A larger (positive) coefficient indicates a greater dip in intensity for the intervening consonant; coefficients with a value of zero indicate no dip in intensity. Expressing the dip in intensity as this coefficient has two advantages. First, it allows for data reduction, expressing dynamic

data as a single parameter. Second, it functions to normalize differences in absolute intensity found across speakers and tokens, due to, for example, distance from the microphone, as all measurements are rendered relative to other samples in the particular acoustic interval under consideration.

ARTICULATORY VOWEL DISSIMILARITY. Images of the tongue recorded with ultrasound were analyzed by generating time-varying articulatory signals with reference to the vowel /a/. In order to prepare the images for the creation of these articulatory signals, each speaker's image set was processed using TRACTUS (Temporally Resolved Articulatory Configuration Tracking of Ultrasound; Carignan 2014). Images were first filtered sequentially to maximize the image variance related to the changing tongue configuration: anisotropic speckle reduction (edge-sensitive noise reduction; Hueber et al. 2007, Yu & Acton 2002), median filtering (localized noise reduction), Gaussian filtering (global noise reduction), and Laplacian filtering (edge contrast enhancement). A region of interest was manually selected around the bounds of tongue movement throughout each speaker's video recording, in order to reduce the amount of variance not related to tongue surface movement. In order to reduce dimensionality, bicubic interpolation was used to resample the filtered image to 30% of the original resolution. The final preprocessed images were then cropped to the region of interest.

The pixels in the preprocessed images were submitted to a principal component analysis (PCA), wherein pixel sites were dimensions and individual image frames were observations. Only ultrasound frames corresponding to the target sequences were included as observations, in order to ensure that each speaker's PCA model accounted for only the variance related to tongue movement that was of interest to our study. Applying PCA in this manner to image sequences (Hueber et al. 2007) not only reduces dimensionality for the reduction of computational costs, but it also allows the resulting orthogonal principal components (PCs) to be used as predictors in further statistical modeling (see below). More detailed information about the treatment of ultrasound frames using this method can be found in Mielke et al. 2017.

For each speaker's PCA model, the number of PCs that were required to explain at least 80% of the total image variance were retained (mean = 25 PCs). The scores for these PCs were then used as independent variables in a linear discriminant analysis (LDA) model trained on ultrasound PCs for the vowel /a/ (i.e. two PC scores for each /...aCa.../ token); the model was subsequently used to predict an LDA score for each frame in each target triphone sequence. Thus, the LDA score for each ultrasound frame across the target /...aCa.../ sequence provides a measure of vowel similarity (i.e. how similar is the tongue shape in each ultrasound frame to the tongue shape used to produce the flanking vowels?). Since the target sequence is flanked by the vowel /a/, LDA similarity is high at these time points. A drop in the LDA score in the middle of the sequence indicates that the articulation diverges from the vowel /a/. For the velar trials, we expect to find a larger drop in LDA score for tokens labeled <k> than for tokens labeled <h>, as stops should be less similar to vowels than approximants are. To provide a compact metric of the dip in LDA score across time, we again used the quadratic coefficient of a second-order polynomial fit to the change in LDA across time. This provides a measure of vowel dissimilarity that is highly comparable to our measure of acoustic intensity change. A larger coefficient indicates greater vowel dissimilarity for the intervening consonant.

4.3. CLUSTERING ANALYSIS. The above phonetic parameters—intensity based on the acoustic signal and vowel dissimilarity based on the articulatory signal—define a two-dimensional phonetic space (see Fig. 2 above). To assess the degree to which the pho-

netic data conforms to categories, we conducted an unsupervised clustering of the data within this two-dimensional space, as well as, for comparison, each phonetic dimension separately. For clustering, we used a Gaussian finite mixture model fitted by the expectation maximization (EM) algorithm using the Mclust package (Fraley & Raftery 2007) in R. Optimal clustering of the data was determined according to the Bayesian information criterion (BIC).

4.4. DURATION. We also measured the phonetic duration of the /aCa/ interval. Because of the similarity of C to the preceding and following /a/ vowels in some tokens, it was sometimes impossible to parse out the C. In contrast to our attempts to segment the velar consonant, C, from the /aCa/ interval, parsing out the entire /aCa/ interval was relatively straightforward. We were able to use standard acoustic diagnostics for the vowels: in the transition to the first /a/, we used an abrupt increase in intensity and corresponding formant structure to segment the start of the /aCa/ interval; correspondingly, we used an abrupt decrease in intensity and loss of formant structure to segment the end of the /aCa/ interval from the following consonant. Our analysis compares the duration of this interval across the dictionary labels and across phonetic categories determined by optimal clustering of the data, as described above.

5. RESULTS. We report the results in four subsections beginning with an analysis of acoustic intensity (§5.1) and vowel dissimilarity (§5.2) independently. These sections compare velar consonants with the palatal stop-approximant contrast. The palatals served as a baseline for demonstrating our methods, since there is no doubt that the palatal stop and the approximant are both phonetically and phonologically distinct. Both velar and palatal consonants are analyzed in terms of the dictionary labels and also via bottom-up clustering based on the phonetic measurements alone. The bottom-up method is particularly important for the velars, given the indeterminacy around the phonetic identity of orthographic <h> and whether there is indeed a phonemic contrast between /k/ and /q/. After reporting independent analyses of acoustic intensity (§5.1) and vowel dissimilarity (§5.2), we report clustering based on a two-dimensional manner space incorporating both phonetic dimensions (§5.3). Finally, we close with an analysis of phonetic duration (§5.4) informed by both the dictionary labels and the categories that emerged from clustering in the two-dimensional phonetic space. Online supplementary materials (accessible at <http://muse.jhu.edu/resolve/104>) include the data measurements from the experiment and an R markdown file with the code used to generate Figs. 7–23.

5.1. INTENSITY DIP.

PHONETIC VALUES BY DICTIONARY LABELS. Figure 7 shows the change in acoustic intensity across the target sequence for palatals (top row) and velars (bottom row). The labels are based on the orthography provided in Table 1. As expected, we observe substantial dips in intensity for consonants relative to the flanking vowels (see also Fig. 1). We can also observe variation in the size of the dips across participants and across consonants. Across participants, there tends to be a larger dip in intensity for stop consonants than for approximants at the same place of articulation. For palatals, the intensity dip is particularly large for participants IL and JC. Both of these speakers show smaller dips for the corresponding velar consonants.

Figure 8 shows the mean quadratic coefficient for polynomials fitted to the change in intensity. Error bars indicate 95% confidence intervals. The left panel shows palatals; the right panel shows velars. We can see that, on average, all speakers differentiate stops from approximants based on intensity dip at both places of articulation. The differences are larger for palatals than for velars, particularly for two speakers, IL and JC.

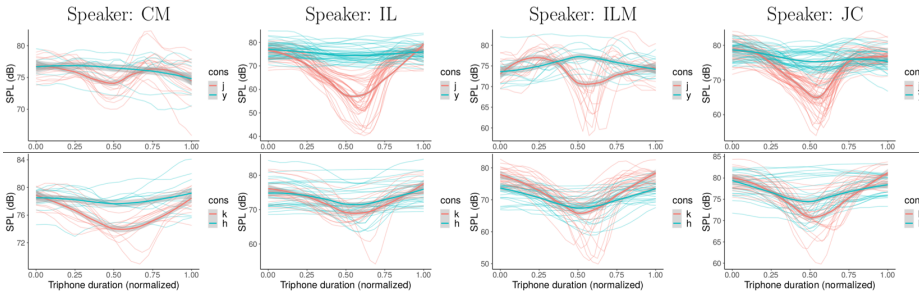


FIGURE 7. Change in intensity across normalized time over [aCa] target sequences. The upper panels show palatals, and the bottom panels show velars. Labels reflect the practical orthography (see Table 1 for IPA mappings).

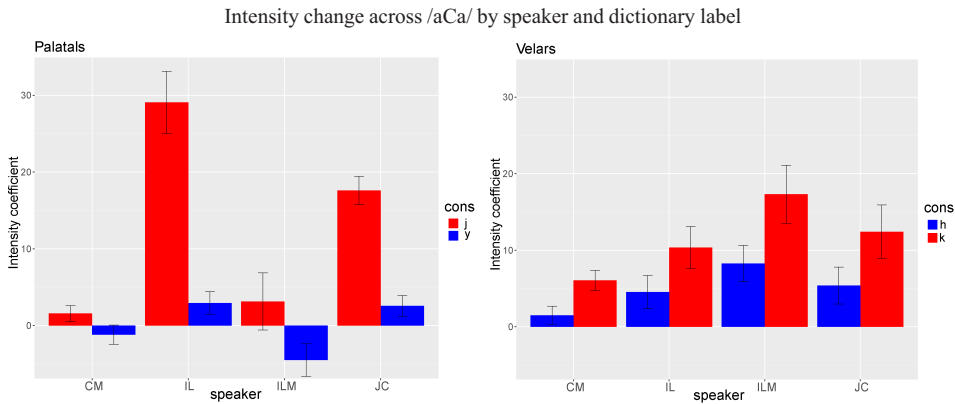


FIGURE 8. Intensity coefficient by dictionary label for palatals (left) and velars (right). Higher values indicate greater intensity dip; zero indicates no intensity dip; negative values indicate an increase in intensity for the target segment. Error bars indicate 95% confidence intervals.

As described in the methods, one advantage of expressing the change in intensity as the quadratic term of a polynomial fit to the data is that this also normalizes for absolute differences in intensity found across recordings and speakers. We also considered further normalizing the quadratic terms within-speaker, but found that this has little effect on the overall distribution.

UNSUPERVISED CLUSTERING OF INTENSITY. Clustering analyses were conducted separately for palatals and velars on the basis of the intensity dip, indexed by the polynomial coefficient. Results are shown in Figure 9. The bottom of each panel shows the values of the intensity coefficient for each token as vertical black lines. Category membership of each token is indicated by a vertical line of a different color corresponding to each black line. For palatals (left), tokens fall within either the blue category (1) or the red category (2). For velars, three categories, shown in blue (1), red (2), and green (3), maximize the likelihood of the data.

To compare the unsupervised classification of the data with the dictionary labels, we imposed an interpretation on each of the categories identified. For palatals, we interpreted the blue category as <y> /j/ because of the smaller intensity dip and the red category as <j> /c/. For the velars, we interpreted the blue category as <a> /a/, the red category as <h> /u/, and the green category as <k> /k/. Given this interpretation, we can evaluate how closely the unsupervised clusters correspond to the dictionary labels

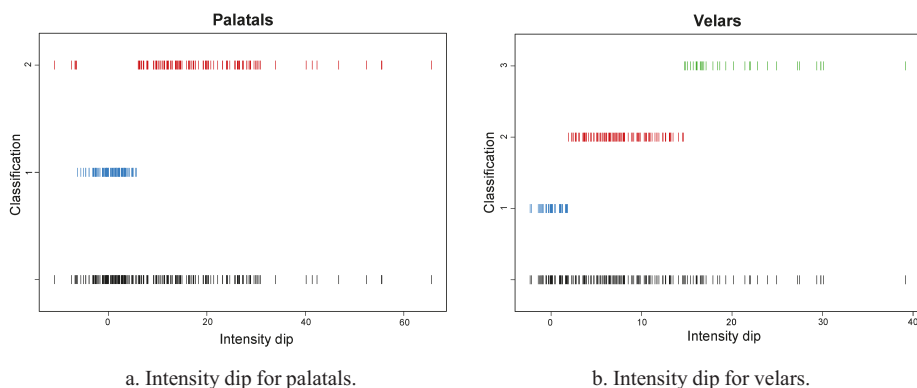


FIGURE 9. Classification results for palatal (a) and velar (b) consonants. The optimal number of categories for palatals was two; that for velars was three.

on a token-by-token basis. This amounts to a kind of ‘accuracy’, if we assume that all of the dictionary transcriptions are accurate in the relevant sense. However, here we take a more conservative approach and refer to this as ‘matching’. The proportion of matches between dictionary labels and the clustering analysis based on intensity dip are given in Figure 10, with the palatals on the left and the velars on the right. For all speakers, the proportion of matches is higher for palatals than for velars. For speaker IL, unsupervised categorization corresponds perfectly to the dictionary labels for <j> /c/. There is also perfect correspondence for speaker CM’s <y> /j/ tokens. For the palatals, most combinations of speaker and consonant are greater than 0.75. These numbers drop sharply for the velars. This is in part because three clusters were found in the unsupervised data for just two segments distinguished in the orthography. Note that for one speaker, CM, the unsupervised clustering based on intensity dip identified no stop consonant at either place of articulation.

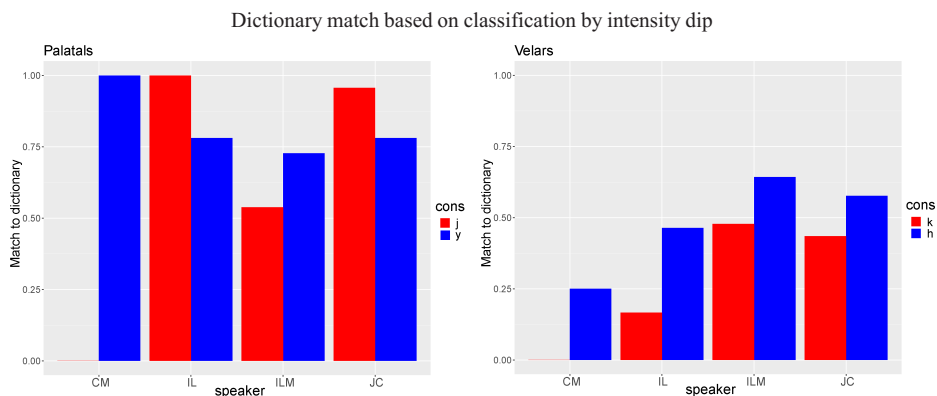


FIGURE 10. Match between unsupervised clusters based on intensity dip and dictionary labels.

Taken together, the analyses in this section show that stops and approximants at both places of articulation are differentiated on average by a dip in intensity, with stops having a greater dip than approximants. Nevertheless, the dip in intensity alone is not sufficient to fully differentiate stops and approximants on a token-by-token basis, and this measure is substantially worse for the velars than for the palatals. We now turn to our

articulatory data to add another angle on possible phonetic differences between <h> and <k> (and /a/).

5.2. VOWEL DISSIMILARITY. Our method of analyzing the articulatory data compares explicitly the similarity of the target /...aCa.../ string to /a/. This method is particularly appropriate in light of the classification results based on intensity. Optimal clustering of the velars resulted in three categories, one of which (the blue category in Fig. 9: right) had a negligible dip in intensity (values around 0) across the sequence. For the purpose of checking the cluster analysis against dictionary transcriptions, we interpreted this category as /a/. The vowel dissimilarity measure allows us to evaluate directly whether the lack of intensity dip in these tokens corresponds as well with tongue shapes similar to /a/.

PHONETIC VALUES BY DICTIONARY LABELS. Figure 11 shows the LDA scores across target /aCa/ sequences, the same sequences used in the intensity analysis. The top row shows palatal consonants; the bottom row shows velars. The high LDA scores at the beginning and end of the target sequences in each panel are due to the similarity of that portion of the signal to the /a/. Several tokens show a dip in LDA score in the middle of the sequence, indicating decreased similarity to /a/. Both of the palatal consonants, <y> and <j>, show dips in LDA scores of similar magnitude. In contrast, some speakers show manner-based differences for the velars. In particular, speaker CM shows a clear dip in vowel similarity for <k> but not for <h>. CM's average for <h> is roughly flat across the sequence, indicating very little departure in articulatory similarity to /a/ across the target sequence. The pattern of vowel dissimilarity for CM corresponds to differences observed in intensity for <h> vs. <k>. The other speakers also showed differences in intensity dip across <h> and <k> that are not as obviously reflected in articulatory differences from /a/. This result can be seen with particular clarity in the quadratic terms of polynomials fit to the change in LDA score across the sequence (Figure 12).

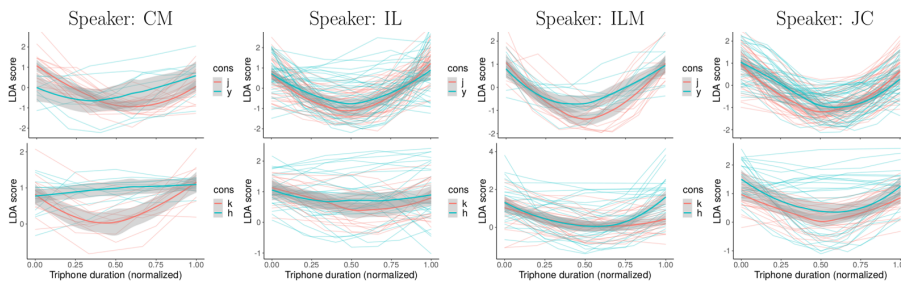


FIGURE 11. LDA scores across time for target sequences. The upper panels show palatals, and the bottom panels show velars. Labels reflect the practical orthography (see Table 1 for IPA mappings).

Figure 12 reports the average quadratic term of polynomials fit to the change in LDA score over time. Larger numbers indicate greater vowel dissimilarity. The palatal consonants (left panel) show high degrees of vowel dissimilarity for both stops and approximants. This is expected, since the place of articulation for the palatals is very different from the flanking /a/ vowels regardless of manner. There is a trend for palatal stops to be even less similar to /a/ than palatal glides, but the overlapping confidence intervals suggest that this trend may not be statistically significant in the current sample. In contrast to the palatals, the velar consonants are differentiated from flanking /a/ to a lesser degree; however, the effect of manner on vowel dissimilarity is greater, at least for one speaker, CM. CM shows a large difference between <k> and <h>, with <h> being highly similar

to vowels and <k> being clearly dissimilar. IL shows a reduced version of the same pattern, while ILM shows the reversed pattern, and JC shows no difference.

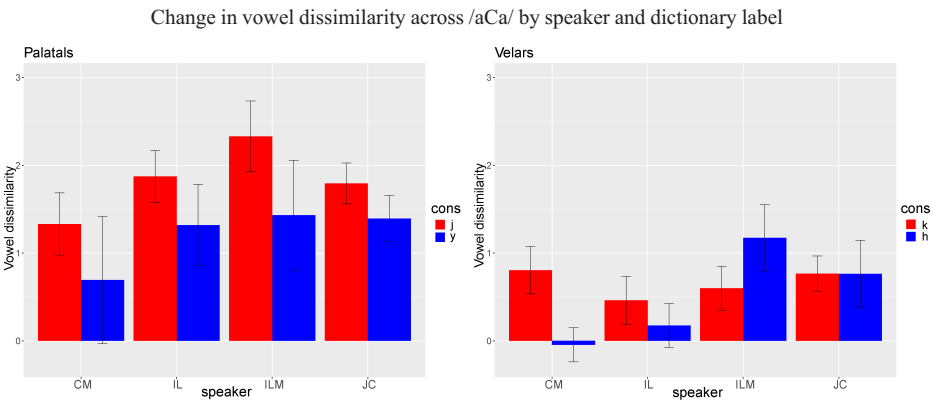


FIGURE 12. Quadratic term of polynomial fit to change in LDA scores. Larger numbers indicate a greater dip in LDA scores, an index of vowel dissimilarity. Smaller numbers indicate tokens that maintained similarity to /a/ across the target VCV sequence. Error bars indicate 95% confidence intervals.

We also considered normalizing the quadratic term for vowel dissimilarity, but just as we found for intensity dip, the distribution of the raw measure was well behaved and not substantially different from the normalized measure.

UNSUPERVISED CLUSTERING BASED ON ARTICULATORY VOWEL DISSIMILARITY. The results of the clustering analysis based on vowel dissimilarity are shown in Figure 13. For the palatals, the optimal fit to the data included just a single category. This indicates that the articulatory data does not contain a reliable enough signal to differentiate between palatal stops and palatal glides on the basis of their degree of dissimilarity from /a/. Both the stops and the glides are different from the flanking /a/ to a similar (and indistinguishable) degree. For velars, however, the vowel dissimilarity measure provides a useful signal. Unlike the intensity dip, for which three categories were discovered, the velars form just two categories in terms of how they differ from /a/. We note, however, that there are many more tokens in the blue category (1), the category that is closer to the vowel, than in the red category (2).

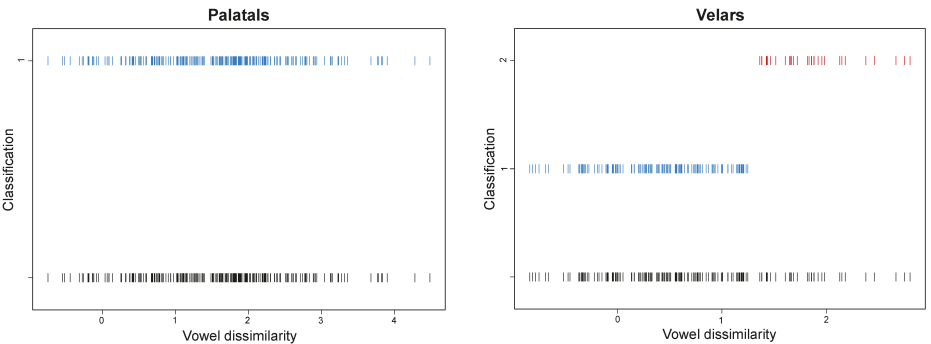


FIGURE 13. Classification results for palatal (left) and velar (right) consonants.

To assess how well the unsupervised clusters map to dictionary labels, we assume that, for velars, the blue category (1) is <h> and the red category (2) is <k>. Since just one cat-

egory was found for the palatals, we cannot really check it against dictionary labels. If we interpret the single category as <y>, then all of the <y>s are matched and <j>s unmatched; if we interpret the single category as <j>, then we get the opposite result.

Figure 14 provides the proportion of matches of the articulatory classifications to the dictionary labels. In general, there were a large number of matches to <h>. For CM, who showed the largest difference between <h> and <k> tokens, there is a perfect match to <h> tokens. Note that the high proportion of matches on <h> and lower proportion of matches on <k> for CM as well as for the other speakers follow from the large number of tokens overall that were classified as belonging to the <h> class, that is, the blue tokens in Fig. 13 (right).

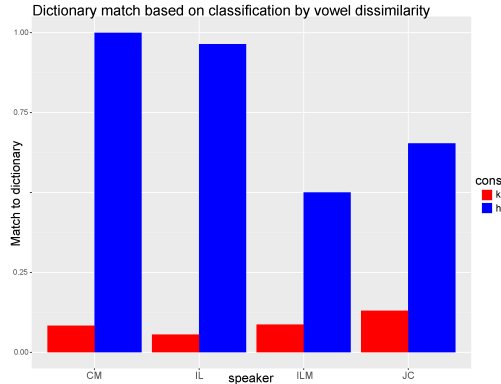


FIGURE 14. Match between unsupervised clusters based on vowel dissimilarity and the dictionary labels for those words.

To sum up the vowel dissimilarity measure, for the palatals, there was a consistent trend across speakers—stops were less similar to the vowels than glides were. However, when viewed through bottom-up classification, vowel dissimilarity was of limited use for the palatals. Palatal stops and glides did not form distinct clusters. This is because both the palatal stop and palatal glide, since they involve a higher and advanced (anterior) tongue body, are very different from /a/. For velars, the vowel dissimilarity measure revealed interesting patterns. Across subjects, the degree to which <k> differed from /a/ was fairly similar, but there was substantial variation for <h>, ranging from the pattern for CM and IL, in which <h> is undifferentiated from /a/, to ILM, who differentiated <h> from /a/ to an even greater degree than she differentiated <k> from /a/. When it came to unsupervised clustering, we found two categories—one large category that overlaps with /a/ (i.e. values of 0 that indicate no change in LDA score across the target sequence) and a smaller one that is differentiated.

5.3. TWO-DIMENSIONAL PHONETIC SPACE FOR MANNER. Having investigated phonetic categories based on intensity dip (acoustic dynamics) and vowel dissimilarity (articulatory dynamics) independently, we now consider how these two dimensions may co-determine phonetic categories.

Before proceeding to the two-dimensional clustering analysis, we first report the correlation between the phonetic parameters that enter into the analysis. Intensity dip has been used as an acoustic proxy for articulatory constriction degree, including in analysis of lenition (Ennever et al. 2017). Although we are unaware of any direct empirical observation of their relation, we find the assumption that reduction of constriction degree would result in an increase in intensity to be reasonable. Since lenition brings the

articulation of a consonant closer to that of a vowel, we expect that our measure of intensity change will be related to some degree to our measure of vowel dissimilarity. Specifically, there should be a positive correlation between these measures—decreases in the degree of intensity dip (indicating lenition) should be correlated with decreases in our measure of vowel dissimilarity. This assumption is evaluated in Figure 15, which plots the polynomial term representing intensity dip (x-axis) against the polynomial term representing vowel dissimilarity (y-axis) for all velar tokens (collapsing over <h> and <k>).

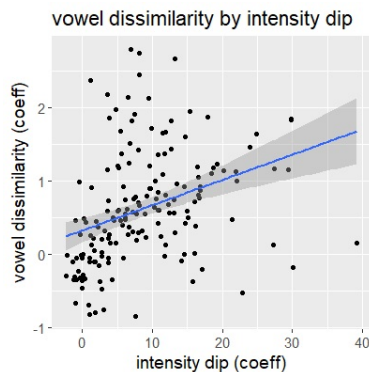


FIGURE 15. Scatterplot of vowel dissimilarity (y-axis) by intensity dip (x-axis) for all velar tokens from all speakers.

Indeed, there is a positive correlation ($r = 0.32$) between vowel dissimilarity (y-axis) and intensity dip (x-axis). Tokens that have a smaller dip in intensity (left side of the figure) tend toward having more vowel-like articulations. There is, however, substantial variation as well, which indicates that articulation (and, specifically, vowel dissimilarity) cannot be inferred from intensity alone. The data includes tokens that have little to no dip in intensity which are nonetheless dissimilar to vowels (top, left quadrant) and tokens that have large dips in intensity that are similar articulatorily to vowels (bottom, right quadrant). The optimal solution for the clustering analysis of tokens shown above in Fig. 15 involved TWO CATEGORIES. Classification results, along with the category centroids, are provided in Figure 16.

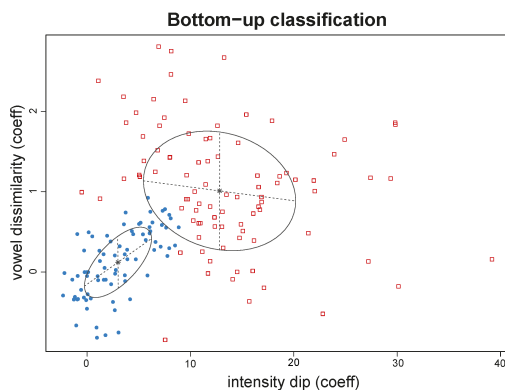


FIGURE 16. Clustering results for velar tokens based on two dimensions, intensity dip (x-axis) and vowel dissimilarity (y-axis).

The two categories found by unsupervised clustering include one in the bottom left quadrant (in blue) and one that forms a diagonal between the top left quadrant and the bottom right quadrant (in red). There are a few tokens in the top right quadrant but not enough to statistically justify a distinct category, based on the data in our corpus. In terms of the phonetic identity of the categories, the blue category overlaps the point of the phonetic space in which both vowel dissimilarity (x-axis) and intensity dip (y-axis) are zero. The zero point indicates when there is NO DIFFERENCE across the entire /aCa/ sequence in vowel dissimilarity or intensity. In short, these tokens are minimally different from the flanking vowel /a/. The red category is different from /a/ in either how the articulation differs from vowels (top left quadrant), the degree of intensity dip (bottom right quadrant), or both (middle of the figure, near the center of the red category).

Looking within the categories, we see now that there is indeed a positive correlation between vowel dissimilarity (y-axis) and intensity (x-axis) within the blue category. Moreover, the positive correlation found within the blue category is substantially stronger ($r = 0.64$) than the correlation across all velar tokens (cf. above $r = 0.32$). For the blue category, a flattening of the intensity contour goes hand in hand with vowel similarity. However, this strong positive correlation is not observed within the red category. For the red category, flatter intensity involves greater vowel dissimilarity. There is actually a negative correlation ($r = -0.27$) between the phonetic parameters within the red category.

When two phonetic parameters dually provide information about a phonetic category, they sometimes show trading relations, such that reduction in distinctiveness on one dimension is compensated by (or trades off with) an increase in distinctiveness on another dimension (e.g. Perkell et al. 1993, Summerfield & Haggard 1977). When it comes to the red category, acoustic intensity and vowel dissimilarity are negatively correlated. This is not expected from the standpoint of acoustic intensity as an index of constriction degree. Although we observe this positive correlation across the data set as a whole and we observe it quite strongly within the blue category, the red category defies physiological naturalness. In the case of the red category, two phonetic dimensions that otherwise tend to be correlated are potentially controlled to maintain distinctiveness. Such decoupling of phonetic parameters is observed in other domains as well, for example, universality of intrinsic f_0 (cf. Connell 2002, Whalen & Levitt 1995). The observation for us reinforces the statistically optimal determination of categories identified by the clustering algorithm. Speakers may be actively maintaining distinctiveness of the red category in one way (vowel dissimilarity) or another (intensity change). While the red category is clearly distinct from the flanking vowels, the same cannot be said for the blue category. Of the two phonetic categories in the data, only one of them (red) is a consonant.

The absence of a phonetic distinction between two consonant categories—there is only a distinction between a consonant (red) and a vowel (blue)—complicates how we evaluate the dictionary labels. For the purpose of assessing dictionary match, we have interpreted the blue category in Fig. 16 as <h>, since it is more similar to a vowel articulatorily and has less of an intensity dip, and the red category as <k>. Figure 17 shows by speaker how dictionary labels are distributed across the two-dimensional phonetic space of our classification. Here we can see that indications of <h> vs. <k> in the dictionary are not particularly good predictors of how individual speakers produce the words. Speaker CM shows the clearest phonetic distinction between <k>-words and <h>-words, but there are still dictionary <k>-words that are classified on the basis of the phonetics as belonging to the vowel-like distribution (blue). CM also has very few

tokens with an intensity dip greater than 10 and no tokens in the upper right quadrant. Speaker JC shows two clear probability masses within each of the dictionary labels, indicating quasi-categorical variation between a more vowel-like and less vowel-like segment within <k>-words and also within <h>-words.

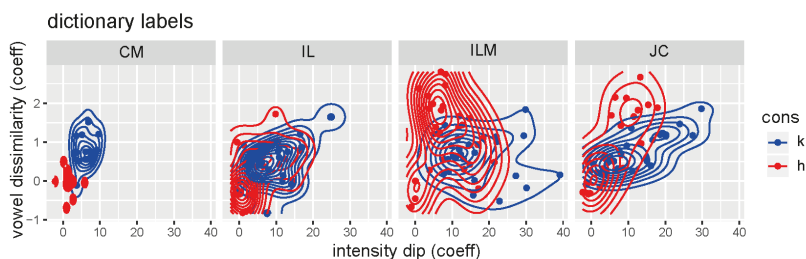


FIGURE 17. Distribution of dictionary labels in phonetic space by speaker.

Bottom-up clustering, illustrated in Figure 18, carves the tokens in a very different way. For JC, the two clusters observed, respectively, within the <k> and <h> groups of the dictionary labels are united in the clustering analysis as one category that is more vocalic (cross-cutting <h> and <k> dictionary labels) and one that is less vocalic. For ILM, we can see that many of the tokens labeled by the dictionary as <h> were phonetically classified as the less vocalic category. ILM in general tended to produce the less vocalic category; CM, in contrast, tended to produce the more vocalic category. CM and IL also had dense clusters, indicating fairly reliable within-speaker variation across the categories. However, categorical phonetic variation did not map onto the phonological differences putatively captured by the orthography. Thus, at least some words had variable realizations across speakers.

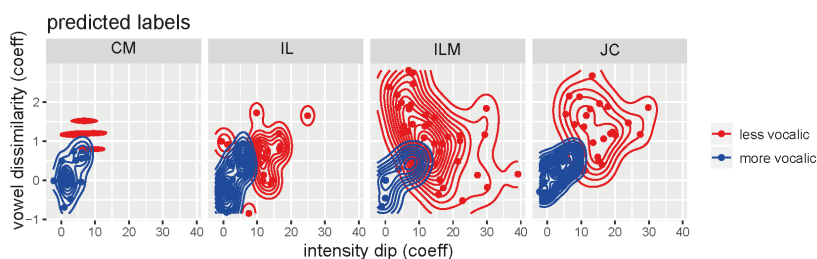


FIGURE 18. Distribution of predicted labels in phonetic space by speaker.

To assess the degree to which specific lexical items showed phonetic variation across subjects, Figure 19 shows the number of tokens from each phonetic category represented by the words in the corpus. All of the words showed categorical phonetic variation in that they were produced as both the less vocalic and the more vocalic categories. The words in the figure are ordered by dictionary label, with <k> on the left side of the figure and <h> on the right. The percentage of tokens produced with the more vocalic category is provided at the bottom of each bar. On average, words written with <h> were more likely (56%) to belong to the more vocalic category than words written with <k> (33%). However, the orthographic distinction masks an impressive degree of phonetic variation found across speakers and items.

We next turn to within-speaker variability. Each row in Figure 20 shows data from a different speaker. The order of the lexical items across the x-axis is the same as in Fig.

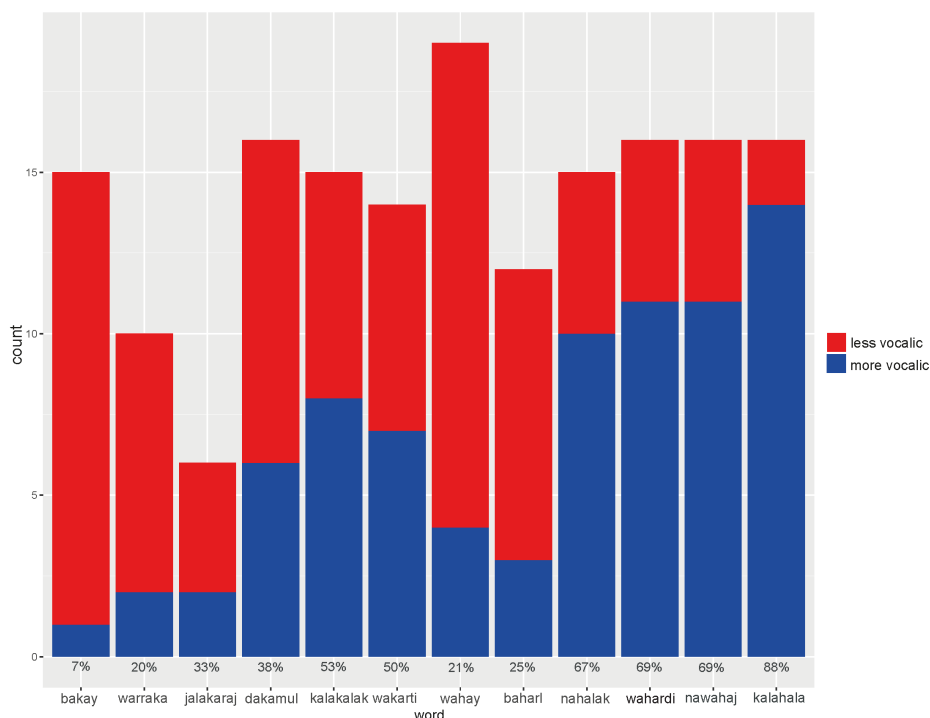


FIGURE 19. The distribution of bottom-up classification categories by word; ‘more vocalic’ refers to the blue category in Fig. 14, and ‘less vocalic’ refers to the red category. Numbers below the bars indicate the percentage of tokens classified as the more vocalic category.

19. As already observed, some speakers tend to produce the less vocalic (ILM) or the more vocalic (CM) variant overall. However, which lexical item is produced with which category varies across speakers. To begin with, consider CM (top panel), who produced exactly two repetitions of each of our target words. Both tokens of <bakay> and both tokens of <kalakalak> were produced with the less vocalic phonetic category (red). The other words were produced with the more vocalic category (blue). Just one word, <wakarti>, was produced with variation by CM, that is, once with the more vocalic category and once with the less vocalic category. Since CM tends to produce the more vocalic variant, we could reason that the words he produced with the less vocalic category would be produced with the less vocalic variant by other speakers as well. However, this is not in fact what we found. The noun <kalakalak> ‘bird’, which CM produced with the less vocalic category (red), was produced exclusively (five times) with the more vocalic category (blue) by JC. Even ILM, who has an overall tendency to produce the less vocalic category (red), produced <kalakalak> once with the more vocalic category (blue). Although CM shows variation for just one of the words, the other three speakers, for which we were able to collect more repetitions, show more variation. Categorical intraspeaker variation was found for four words for IL, four words for ILM, and two words for JC. The words that show variation are also not consistent across speakers. There is no word that is produced consistently with the same category across speakers. The ones that come closest are <bakay>, which shows a strong tendency to be produced with the less vocalic category (red), and <kalahala>, which shows the strongest tendency to be produced with the more vocalic category (blue).

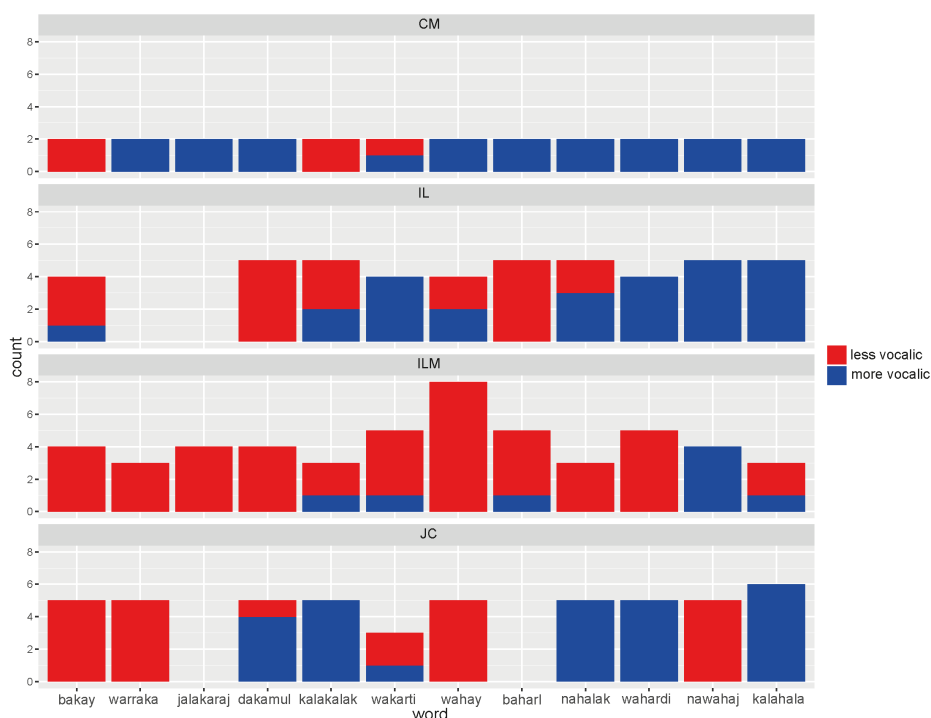


FIGURE 20. The distribution of bottom-up classification categories by word and by speaker; 'more vocalic' refers to the blue category in Fig. 14, and 'less vocalic' refers to the red category.

In sum, all lexical items in our corpus vary in their realization across the two phonetic categories, either within-speaker, across-speaker, or both. Thus, while there is some statistical evidence for two distinct categories, we have no evidence that these two categories are used consistently with particular lexical items by the community. Given the variation found across speakers, it seems unlikely that the phonetic distinction between categories is useful in distinguishing words, unless the identity of the speaker (and hence the speaker-specific use of the distinction) is also known.

5.4. SEGMENT DURATION. Lastly, we report the duration of the trigram VCV sequences. Figure 21 reports trigram duration by dictionary label. There is no significant difference in duration between trigram sequences labeled in the dictionary as <h> and items labeled as <k>. However, our previous analyses indicate that some words were not produced according to the dictionary forms, at least not all the time.

Figure 22 shows trigram duration according to the unsupervised category labels of §5.3 (using both phonetic parameters). Here, we see differences in duration depending on how the token was classified, especially for JC (but all subjects pattern in the same direction). Duration differences emerge from classification based on intensity dip and vowel dissimilarity. By corollary, it follows that the combination of these two phonetic parameters predicts duration. Each factor is correlated with duration independently (duration and acoustic intensity: $r = 0.66$; duration and vowel dissimilarity: $r = 0.36$), and they show additive effects in predicting duration together. This was established by fitting a series of nested linear mixed-effects models to trigram duration. The baseline model contained random intercepts for speaker and lexical item. We also evaluated the maximal random-effects structure, with random slopes for all fixed factors, but the in-

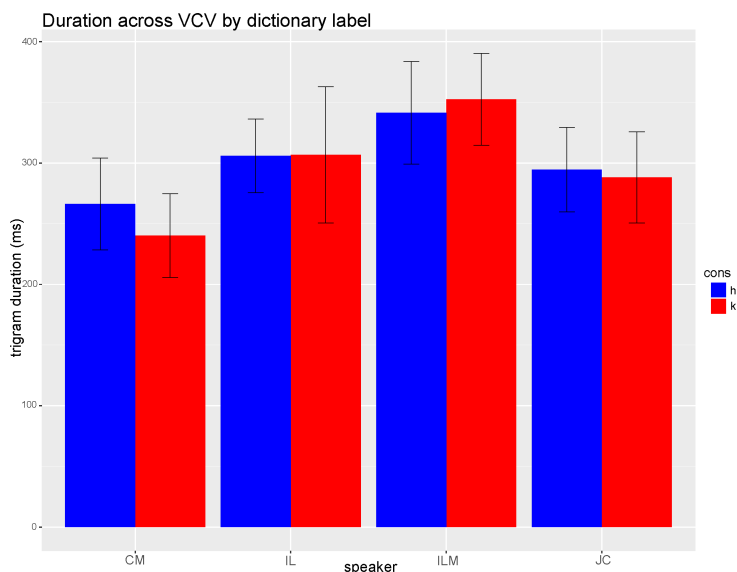


FIGURE 21. Trigram duration according to dictionary labels. There is no reliable difference between <h> and <k>.

clusion of random slopes was not justified in that the random slopes did not lower the Akaike information criterion (AIC). To the baseline model (which included only random effects), we proceeded by adding fixed factors one at a time and comparing the more complex model to the less complex model. We first added intensity dip as a fixed factor to the model, which showed significant improvement ($\chi^2 = 42.67$, $p < 0.01$) and justified the increased model complexity, as indicated by a decrease in AIC from -478.16 (baseline) to -518.83 (intensity dip). Adding vowel dissimilarity resulted in further improvement ($\chi^2 = 4.46$, $p < 0.05$, AIC = -521.28). These results show that, consistent with Fig. 22, duration can be viewed as codetermined by both intensity dip and vowel dissimilarity. We also tested for an interaction term, intensity dip * vowel dissimilarity, but found that the interaction did not improve the model above and beyond the additive effects (intensity dip + vowel dissimilarity). This indicates that the two phonetic parameters, acoustic intensity dip and vowel dissimilarity, make independent contributions to trigram duration.⁷

The entire distribution of duration values in Iwaidja is provided in boxplots in Figure 23. The label [aaa] refers to trigrams classified as belonging to the more vocalic category; the label [auja] refers to trigrams classified as belonging to the more consonantal category; the label [á] is the duration of stressed singletons. A linear mixed-effects model of duration, with subjects and items as random effects, confirms the trend shown in Fig. 23. The [aaa] sequences with lenited consonants are more than TWICE the duration of a stressed [á] ($t = 7.35$, $p < 0.001$). This indicates that the phonetic signal maintains substantial temporal duration, even for tokens involved in complete lenition based on our other acoustic and articulatory measures.

⁷ We also conducted unsupervised clustering based on three phonetic parameters, including duration along with intensity dip and articulatory dissimilarity. Including duration resulted in the same general pattern of results, but slightly lowered the % match to dictionary labels.

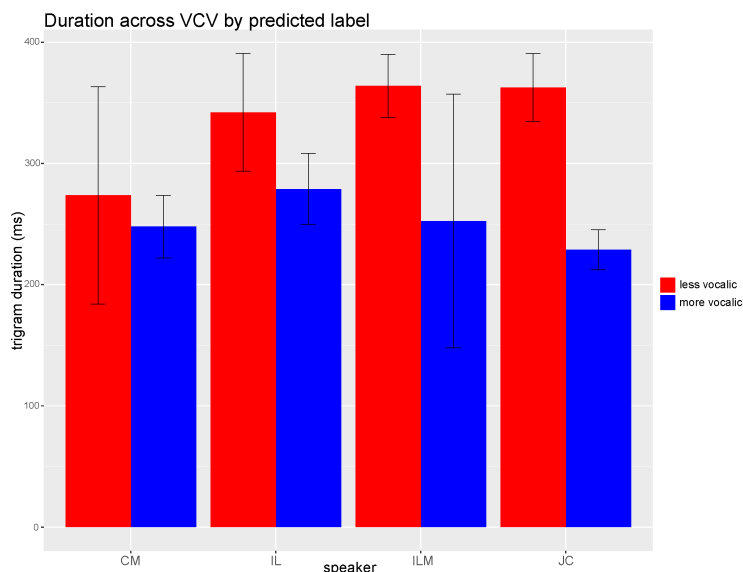


FIGURE 22. Trigram duration according to classification label. The less vocalic of the categories is longer in duration than the more vocalic.

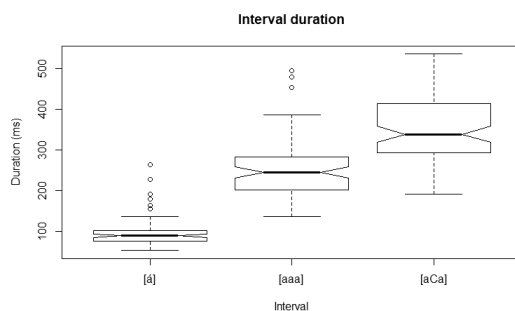


FIGURE 23. Duration of the 'more vocalic' trigrams ([aaa]), the 'less vocalic' trigrams ([aCa]), and the stressed singleton vowels ([ā]).

From the standpoint of the undershoot hypothesis (see §2.2), the more likely direction of causation is that reduction in duration influences both intensity dip and vowel dissimilarity. Despite the reduction, however, it is still the case that the lenited (more vocalic) variants remain more than twice as long as singleton vowels. Iwaidja does not have phonemic vowel length and it does not permit hiatus, so there is no natural long-vowel baseline against which we can evaluate the resulting interval. The average duration of singleton /a/ in a stressed syllable in Iwaidja is 95 ms. The average duration of trigram sequences classified as vocalic (lacking a change in intensity and articulatory difference from /a/) is 252 ms, 62 ms greater than two times the average singleton /a/. As a point of reference from a language with a three-way vowel length contrast, Dinka, Remijsen (2013) reports the average duration of 'overlong' vowels to be 194 ms (cf. short vowels = 70 ms, long vowels = 113 ms). All of this suggests that the [aaa] interval is longer than the combination of the flanking /a/ vowels. Rather, it seems that some of the temporal duration of the (missing) velar contributes to the total duration of the interval.

To summarize the duration results, there was no significant difference in the temporal duration of intervals corresponding to <aha> and <aka> in the orthography. However, there were significant duration differences between the phonetic categories obtained by bottom-up clustering of tokens: [aua] was longer than [aaa]. Reduction of the velar stop goes together with some reduction of phonetic duration. However, the [aaa] interval remained significantly greater than two times the duration of stressed vowels. Outside of intervocalic lenition, vocalic intervals of this duration are nonexistent in the language (i.e. no vowel length contrasts, or vowel sequences).

6. DISCUSSION.

6.1. PHONETIC CATEGORIES. We analyzed audio and ultrasound recordings of words expected on the basis of past documentary work to contain velar consonants. Bottom-up classification of velar tokens in a two-dimensional manner space (intensity dip by vowel dissimilarity: Fig. 2) revealed two phonetic categories: one category overlaps in our two-dimensional manner space with [a]; the other is more consonantal, either on the acoustic parameter (intensity dip) or the articulatory parameter (vowel dissimilarity), or both, and can be labeled [u].

Distributional analysis of phonetic parameters is widely assumed to play an important role in the formation of linguistic categories in acquisition (e.g. Maye et al. 2008, McMurray et al. 2009, Pierrehumbert 2003). We find it promising that a purely statistical method of classification identified exactly the number of phonetic categories (two) that were expected on the basis of past documentary work. However, we do not wish to imply that our specific method or, in a more general sense, any purely statistical method of phonetic categorization necessarily reveals the number and nature of allophones in a data set. Since statistical separation reveals two categories in our data, it seems reasonable to evaluate whether separation in phonetic space corresponds as well to the proposed phonemic contrast.

What is most clear from the data is that there is substantial variation in the phonetic realization of nonnasal velar consonants. In the discussion that follows, we continue to refer to the consonantal (or ‘less vocalic’) category that emerged from the analysis as [u] and the vocalic (or ‘more vocalic’ category) as [a]. These two categories represent the optimal statistical solution to phonetic categorization, given our criterion (BIC). These two categories may not be allophones in the standard sense—the [a] category consists of a tight cluster of tokens, while the [u] seems to function more like an ‘elsewhere’ category, in that it subsumes a wide range of phonetic variation. In evaluating the distribution of these categories across words and speakers, we assess the hypothesis that phonetic categories based on statistical separability subserve a hypothesized phonemic contrast. We do so in light of the mixed evidence for this contrast based on standard noninstrumental methods (see §2.1).

6.2. SPEAKER-SPECIFIC MAPPINGS TO LEXICAL ITEMS. The frequency of the two phonetic variants differed across words and also across speakers. All speakers produced each of the phonetic variants some of the time, and none of the words in our corpus were consistently produced with just one of the phonetic variants (Fig. 20). Figure 24 below exemplifies speaker variation for one word, <kalakalak> ‘bird’, against the backdrop of the two phonetic categories in the data. The thick lines show probability distributions over vowel dissimilarity for the phonetic categories. The dark blue line, labeled ‘more vocalic’, shows the vowel-like category; the red line shows the consonantal (i.e. ‘less vocalic’) category. There is a reasonable degree of separation between these categories on

the vowel dissimilarity dimension.⁸ The dotted lines show distributions over this same phonetic space for one word, <kalakalak> ‘bird’, as produced by three speakers. Speaker CM’s productions of this word are well within the red category (less vocalic), speaker JC’s productions fall within the blue category (more vocalic), and speaker IL’s productions fall in the middle. It is also possible to read this figure as speaker-specific degrees of lenition, with the greater constrictions found on the right (greater vowel dissimilarity) and the weaker constrictions found on the left.

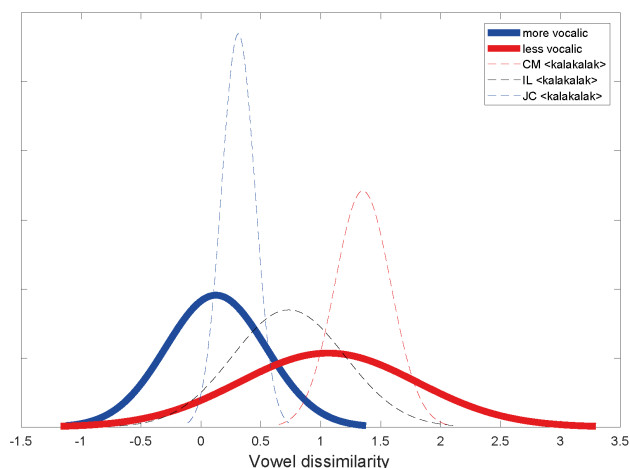


FIGURE 24. Probability distribution functions over the vowel dissimilarity parameter (based on ultrasound analysis). The thick lines show the distribution for the phonetic categories (blue = more vocalic, red = less vocalic). The dotted lines show speaker-specific distributions for the word <kalakalak> ‘bird’. Speaker CM’s pronunciations fall largely within the less vocalic (consonantal) range, speaker JC is in the more vocalic range, and speaker IL is in between.

Figure 24 exemplifies continuous variation within the phonetic manner space for a single word which transgresses the two phonetic categories. Expressed in terms of categories, the phonetic category used for any given word varies by speaker (see also Fig. 20). It is also the case that the phonetic category used by a particular speaker varies by word. That is, all speakers use both categories in at least some words. For example, CM always produced <kalakalak> with the consonant category (i.e. less vocalic) and <warraka> ‘water lily’ with the vowel category (i.e. more vocalic). JC was also internally consistent for these words, but the pattern was reversed: JC produced <kalakalak> with the vowel variant and <warraka> with the consonant variant (Fig. 20). Other words were produced variably with both categories by the same speaker. All speakers produced at least some words consistently with the blue category, some with the red category, and some that were variable (either blue or red).

It is clear from these results that the two phonetic categories are not distributed consistently across the lexicon in the same way across speakers. That is, which word is produced with which phonetic category does not constitute shared knowledge of our speakers, at least not in any absolute (i.e. deterministic) terms. Because of this, to use the phonetics of the nonnasal velar consonant to contribute to word recognition, it is

⁸ For simplicity of visual display, we show just one of the two phonetic dimensions, but the main concept illustrated in Fig. 24 generalizes to the two-dimensional space including vowel dissimilarity and intensity dip.

necessary to know the speaker-specific distribution of the phonetic categories. In this respect, the phonetic variation has the functional equivalence of talker-specific phonetic patterns that can be adapted to on a talker-by-talker basis (Kraljic & Samuel 2006, Norris et al. 2003, Samuel & Kraljic 2009, Sumner & Samuel 2009) but not across a group of talkers (Eisner & McQueen 2005, Shaw et al. 2018).

6.3. DICTIONARY LABELS. When it comes to documenting the lexicon of a language, how should the type of variation found in Iwaidja velars be represented? The working dictionary of Iwaidja has transcribed a difference between velar consonants that is not reliable across speakers. The dictionary labels did, however, capture tendencies for words to occur with one variant or another. The words in our study written with <h> in the dictionary were produced with the more vocalic variant, [a], between 21% and 88% of the time (average 56%), while the words written with <k> were produced with the more vocalic variant between just 7% and 53% of the time (average 33%). In this way, the orthographic labels seemingly discretize a gradient tendency for a word to be more or less vocalic, as averaged across our speakers.

At the level of individual speakers, however, no single speaker shares the velar distinction as represented in the dictionary. The speaker that comes the closest is CM. This was the only speaker who maintained roughly distinct pronunciations for <h> and <k> (Fig. 17). CM was born around the same time as ILM. However, CM and ILM show very different patterns (in which variants are used with which words). It is somewhat ironic that CM is the speaker that matches the dictionary labels most closely, since this speaker has the highest degree of lenition overall. All of CM's velar productions occupied the lower left quadrant (vowels) of the manner space (Figs. 17, 18). More of CM's tokens were classified as vocalic than of any other speaker (Fig. 18), and even the small number of tokens that were classified as consonantal were close to the vocalic category boundary. Note that CM has also generalized intervocalic lenition to the palatal place of articulation; he is the only speaker whose bottom-up classification of the palatals based on intensity dip did not produce a close match to the dictionary. Speakers ILM and JC produced many more consonantal variants of the velar than CM (Fig. 20) and used a greater range of the consonant manner space, that is, more tokens in the upper right, upper left, and lower right quadrants. However, the patterns of phonetic variants for these speakers differed from CM, from each other, and from the dictionary (Fig. 17). Given the range of interspeaker variation, there is no way to select which words should be written with <k> and which with <h>. The words written with <h> vs. <k> on the basis of one speaker would not generalize to the next. To capture variation across speakers, it would therefore be necessary to list both <k> and <h> as variants in intercontinuant position for every word in our corpus. Essentially, in intercontinuant position <h> is always an option, since this is the environment of lenition. Rather than having all words written with both <k> and <h> variants, a simpler solution would be to represent all words with just one letter. As there is something akin to free variation, there is little to lose in collapsing the distinction to describe lexical patterns shared across speakers.

6.4. SYNCHRONIC CONTRAST WITHIN THE NONNASAL VELARS. In canonical cases, phonemic contrast meets the following three criteria: (i) a phonetic difference between two sounds that (ii) map to distinct sets of lexical items, including minimal pairs, and (iii) speakers of the language use the same distribution of phonetic categories across words. However, oftentimes contrast is not so clear cut. Any one of the three criteria may fail. The phonetic difference may be reduced but not obliterated, as in the case of incomplete neutralization (Braver 2014, Warner et al. 2004). The sounds may contrast

in only a subset of the words in which they occur, that is, marginal contrasts (Hall 2009). For example, until recently the Japanese coronal contrast between alveolar stops and alveopalatal affricates was maintained only before vowels /a/, /u/, /o/, being neutralized before /i/, where only the alveopalatal occurred, and before /e/, where only the alveolar occurred (Shaw & Balusu 2010). The contrast is marginal, in the sense of Hall (2009), because it occurs only in some environments. The case of German [s] ~ [z] is even more striking. The distribution is almost completely predictable, but there exists one minimal pair in the regular lexicon: *reisen* [z] ‘travel’ vs. *reißen* [s] ‘tear’.⁹ Finally, there are cases in which speakers differ in which words use which sounds. For example, English pronunciation variants for *data*, [dætə] vs. [dēɪtə], can vary by speaker, with some speakers producing both. Crucially, however, a similar alternation between [æ] and [ē] is not permitted for all English words: *bat* [bæt] and *bait* [bēɪt] remain contrastive even though the vowels are interchangeable in *data*.

For the Iwaidja nonnasal velars, there is also one minimal pair that we are aware of, given in 1 (§2.1). Although we do not have ultrasound data on this pair, we were able to consult three native speakers, who are interested in revitalization efforts involving teaching Iwaidja to nonnative speakers. In the context of a discussion of possible pronunciations of the word, three speakers agreed that [ɟaga] is an acceptable pronunciation of both <raka> ‘he threw it’ and also of <raha> ‘it is upside down’. However, the speakers also agreed that [ɟauɟa], [ɟaaa], and [ɟaa] are acceptable pronunciations of <raha>, but that only [ɟaga] and [ɟaka] are acceptable forms of <raka>. In this case, the word written with <h> allows for a wider range of pronunciations than the word written with <k>, which does not seem to allow vocalic pronunciations. We obtained judgment results for other words in our corpus as well. Although this is a very different kind of data from what is the primary focus of the article, it does not provide conflicting evidence. The nonnasal velars permit a wide range of phonetic variation. The range of variation for specific words is narrower than the entire range. In production, specific speakers sample differently from the range of possible variants.

Outside of the intercontinuant environment, there is no contrast at all between nonnasal velars. In the dictionary, there is only <k>, which surfaces variably as [k] and [g] in utterance-initial and utterance-final positions and as [g] in the environments following nasals and preceding stops. Since there is no voicing contrast elsewhere in the language, this phoneme has been analyzed as /k/ (e.g. Table 1). In the intervocalic environment, nonnasal velars take on a wide range of phonetic realizations, some of which have been transcribed as <h>, in phonemic opposition with /k/. Our analysis found that the optimal statistical clustering of the data involves two phonetic categories in the intercontinuant environment. This establishes the first of the three criteria for phonemic contrast listed at the start of this section. However, the other two criteria are not met. The contrast is marginal at best, since it is only in intercontinuant environments that the two phonetic categories are found. The third criterion is also not met, as there is little interspeaker agreement on which words contain which variant. It is as if the cross-speaker variation in [æ] vs. [ē] for the English word *data* extended across the lexicon so that, for example, *bait* also varied between [bēɪt] and [bæt] on a speaker-by-speaker basis.

The alternative to positing a phonemic contrast is to treat all nonnasal velars as belonging to the same phonemic category. Outside of the intercontinuant environment,

⁹ If place names are added, then there is another one: *Meißen* [s], the city in Saxony, vs. *Meisen* [z], the plural of *Meise* ‘tit (bird species)’.

nonnasal velars typically take the form [g] or [k] and have been analyzed as /k/. In intercontinuant position, realizations range from [g] through the manner space from [uɟ] to [a]. The evidence suggests that this range of phonetic variation does not support a phonemic distinction, even a limited one, that is, one appearing in intercontinuant position only. When the range of continuous variation, corresponding to gradient degrees of lenition, was operationalized in terms of the statistically optimal categories, we found two phonetic categories, [a] and [uɟ]. These categories were variably distributed across lexical items and speakers. The range of continuous variation that the categories represent is presumably sensitive to other factors that influence lenition, such as frequency, word length, functional load, and speech rate, although a detailed examination of this is beyond the scope of this article. On this plausible alternative, instead of having two contrastive categories, Iwaidja has one nonnasal velar consonant which variably lenites such that it can fall within [uɟ] or [a]. Per the caveats mentioned at the outset of the discussion, it is not entirely clear that [uɟ] is an allophone in the traditional sense, as it shows substantial internal phonetic variation. The data is fully consistent with a single nonnasal velar phoneme that lenites to various degrees when it occurs between continuants.

6.5. IWAIJDJA LENITION IN THE BROADER TYPOLOGICAL CONTEXT. In light of the results summarized above, we can say more about the nature of lenition in Iwaidja, as it relates to crosslinguistic patterns. Since the evidence favors a lack of phonemic contrast between /uɟ/ and /k/, we can appropriately characterize intervocalic velar lenition in Iwaidja as nonneutralizing. This brings the description of Iwaidja lenition in line with what Katz (2016) described as continuity lenition—at least insofar as (i) contrast is preserved and (ii) the continuity of the acoustic intensity profile within the prosodic word is improved. Continuity lenition, according to Katz (2016), has the function of highlighting prosodic boundaries, the same function as fortition processes that target the edges of prosodic domains. This characterization is particularly intriguing for Iwaidja in light of how the temporal duration of words is largely preserved, even as the velar gesture is reduced to [a].

The formal framework that Katz proposes involves BOUNDARY DISRUPTION constraints, which function as markedness constraints in OPTIMALITY THEORY (Prince & Smolensky 2004). Boundary disruption constraints penalize decreases in intensity of a specified magnitude and duration that do not occur at word boundaries. The velar lenition pattern in Iwaidja is likely consistent with this type of markedness pressure, since lenition tends to result in higher acoustic intensity levels within word boundaries (or, more precisely, reduction of the constituent-internal decrease in intensity). Moreover, preservation of timing is also conceptually consistent with the functional motivation proposed by Katz. If continuity lenition functions to enhance prosodic boundaries, it should be unsurprising that it also preserves the temporal structure of words. Listeners are known to attune closely to the details of timing to make decisions about word boundaries (e.g. Cutler & Butterfield 1992, Cutler & Clifton 1984), and disrupting timing can lead to systematic misparsing (e.g. Dille & McAuley 2008). Given these facts, and the insights of Katz (2016) and Kingston (2008) regarding continuity lenition, we might expect temporal preservation to be the norm for continuity lenition, particularly cases, such as Iwaidja, that (variably) lenite to a vowel.¹⁰

¹⁰ A referee suggested that /d/ lenition in Caribbean Spanish may be another example of temporal preservation following lenition.

Alongside the pattern of continuity lenition involving velars, Iwaidja also has a number of lenition patterns better characterized as loss lenition, in Katz's (2016) terms. These patterns, described in the appendix, differ from the intervocalic lenition of velars in that they are both contrast-neutralizing and interact with the morphology in complex ways. Interestingly, loss lenition in Iwaidja does not involve alternation between [k] and [u], suggesting that the continuity lenition observed for velars may be distinct from loss lenition affecting other places of articulation.

Lenition in intervocalic position is common crosslinguistically (Kirchner 2001), and particularly so for velars. In Danish, intervocalic weakening of velars preceded diachronically the lenition of labials in the same environment (Foley 1970:90). In Finnish, intervocalic velars in the nominative (strong form) alternate with zero in the genitive (weak form)—for example, [sika] 'pig.NOM' vs. [sian] 'pig.GEN'—while, in the same environments, /t/ and /p/ are weakened but not deleted (Skousen 1975). In the Adama dialect of Fula, /k/ lenition is also more extreme than other places of articulation: /k/ reduces to /h/, while in the same environments /t/ is preserved and /p/ reduces to /f/ (Hyman 1975:167). In Indonesian, /k/ reduces to [ʔ] in coda position, while the place features of /t/ and /p/ are preserved (Soderberg & Olson 2008). Other cases of intervocalic velar lenition include Baale (Yigezu & Dimmendaal 1998), Insular Spanish (Oftedal 1985), Andalusian Spanish (Romero Gallego 1996), Tatar (Poppe 1963), West Tarangan (Nivens 1992), and Modern Welsh (Halle & Clements 1983).

The crosslinguistic tendency for velars to lead lenition (and continuity lenition in particular) may have many sources. Since few languages have velar approximants (as phonemes), there is seldom pressure to maintain closure to preserve meaning. Another factor is the sheer bulk of the tongue body as an active articulator (cf. tongue tip, tongue blade, lips), the distance to the soft palate, and the force required to maintain complete closure. Aerodynamic concerns also conspire against complete velar closure, since the cavity behind the constriction is smaller for velars than for more anterior constrictions.

There could also be factors related to the temporal control of articulators involved in the intervocalic lenition of velars. Both velar stops and vowels have the tongue body as primary articulator. When overlapped in time, vowel gestures and dorsal consonants place competing demands on the movement of the tongue body. The gesture for a velar stop dictates a constriction (constriction degree: closed), while the temporally overlapped vowel dictates an opening (constriction degree: wide). Such a conflict can be resolved by compromising on the competing production goals, for example, 'gestural blending' in the ARTICULATORY PHONOLOGY framework (Saltzman & Munhall 1989).

Temporal control of gestures is known to be language-specific (e.g. Hermes et al. 2017, Shaw & Gafos 2015), as are the parameters that dictate gesture-blending outcomes (e.g. Iskarous et al. 2012). Preferential lenition of velars (as opposed to coronal or labial consonants) would follow from certain combinations of timing control and blending parameters. For example, timing vowels such that they overlap with the intervocalic consonant, as has been proposed for Italian (Smith 1995) as well as for some languages with vowel harmony (Benus & Gafos 2007, Gick et al. 2006, Gordon 1999, Smith 2018), together with blending parameters that favor vocalic control of the tongue body (as opposed to consonantal control), would yield velar lenition. An interesting prediction of this account is that there would also be temporal preservation of the larger VCV interval, as we have observed in Iwaidja. The reason is that, under this scenario, lenition follows from spatial reduction, due to blending, instead of temporal reduction.

The possibility of spatial reduction as independent of temporal reduction was explored extensively by Lavoie (2001), who argued that reduction in temporal duration can make

a voiceless stop sound more voiced, leading to, for example, /k/ → [g] lenition, while reduction in gestural magnitude leads to more open constrictions, such as /k/ → /x/ lenition (Lavoie 2001:21). More recently, Cohen Priva and Gleason (2020) put forward the hypothesis that all lenition may have its basis in temporal shortening (temporal shortening itself often a consequence of informativity) leading to spatial reduction (see also Bauer 2008). We argued above that temporal preservation is consistent with the functional motivation behind the Kingston-Katz proposal. However, strictly speaking, the Iwaidja results are not incompatible with Cohen Priva & Gleason 2020, as there is indeed some temporal shortening accompanying lenition: [aaa] is shorter than [auqa] (Fig. 21). Whether the degree of shortening is sufficient to predict the degree of spatial reduction observed in our data requires modeling work beyond the scope of this article.

Overall, it seems plausible to us that temporal reduction of low-information segments (Cohen Priva 2017) and maintenance of prosodic boundaries (Katz 2016) are both active functional pressures at work in lenition patterns. How these pressures interact with each other, particularly when they conflict, and how they correspond to the details of articulation at the level of gestural coordination are important questions for future research. For the case of Iwaidja, a crucial step forward has been determining the phonemic status of the velars among the phonetic variation conditioned by lenition.

7. CONCLUSION. Past work on Iwaidja has argued for a limited contrast between a velar stop and a velar approximant, occurring only in intercontinuant position, the environment of consonant lenition. Our analysis of combined ultrasound and acoustic data revealed a wide range of phonetic variation in this context and casts doubt on the phonemic contrast. We showed that phonetic categories, determined by bottom-up classification of the data, do not consistently map to the same lexical items across speakers. In the absence of native-speaker agreement on which words contain which phonetic variants, we put forth the alternative analysis that there is only one nonnasal velar consonant category, which, in intervocalic position, varies along a gradient cline from [u] to [a]. Individual speakers differ in the degree of lenition used in particular words and hence which phonetic category best describes the surface form. In reaching this conclusion, we developed new methods for analyzing field-based ultrasound and for quantifying gradient variation in manner within a two-dimensional phonetic space. The [a] variant of the nonnasal velar has the same tongue shape as /a/ vowels found elsewhere in the language but maintains its temporal extent, for example, /...aka.../ → [...aaa...]. This pattern, including the range of phonetic realizations ([u] to [a]), and the maintenance of temporal duration are consistent with the Kingston-Katz proposal for continuity lenition, as lenition both preserves contrast and reduces the intensity drop within the prosodic word.

APPENDIX: LOSS LENITION IN IWAIJDJA

In addition to the case of continuity lenition that is the main focus of the article, Iwaidja shows both diachronic and synchronic evidence for the operation of loss lenition processes. Of relevance to this article is the fact that velars show very limited participation in loss lenition processes. In this appendix, we examine the diachronic patterns and then turn to consider synchronic morphophonemic alternations that may also be characterized as loss lenition.

One possible hypothesis about the velars is that there was historically a contrast between *k and *uq in Iwaidja and that synchronically this contrast is in the process of being neutralized. Could there have once been a stage of pre-Iwaidja where the distribution of [g] vs. [uq, a] realizations showed inter- and intraspeaker consistency across the lexicon? If so, we would have to view the synchronic facts as resulting from neutralization of this pre-Iwaidja distinction. To evaluate the plausibility of this scenario, we consider correspondences between Iwaidja and related languages.

Iwaidja is a member of the Iwaidjan family, whose other confirmed members are Amurdak, Garig, Ilgar, and Mawng (Mailhammer & Harvey 2018). Mailhammer and Harvey (2018:341–47, 54–55) reconstruct an opposition between *k and *tʃ for Proto-Iwaidjan. However, this reconstruction is based on and is entirely dependent on there being a well-supported synchronic opposition between /k/ and /tʃ/ in the daughter languages. Current analyses of Mawng propose an opposition between /k/ and a fricative /ɣ/ (Capell & Hinch 1970:16, Singer 2006:11). There is currently no detailed phonetic research on velars in Mawng, and the exact realization range of the putative /ɣ/ segment is unknown. As discussed in §2.1, the putative Iwaidja opposition has also been characterized as a stop vs. fricative opposition (Pym & Larrimore 1979:3–4). However, there were no fricative realizations of velars in our Iwaidja data.

One of the earliest accounts of Iwaidja, a sketch grammar by Capell, notes that the contrasts between the velars in Iwaidja is unclear, with no minimal pairs; however, the contrast was posited because it apparently exists in Mawng (Capell 1962:152). The proposed Mawng opposition has the same distribution as the proposed opposition between /k/ and /tʃ/ in Iwaidja. It is found only in the environment [+cont] _ [+cont]. Garig and Ilgar are now extinct, but there are substantial materials that show significant congruence with Iwaidja, including the distribution of the hypothesized velar opposition. Our data shows that there is substantial variation in how speakers produce the velar consonant in Iwaidja and that the observed variation does not map onto the putative opposition very closely. Given the essential isomorphy of the proposed oppositions in Mawng and Iwaidja, and the absence of phonetic data in Mawng, Ilgar, and Garig, much of this broader debate hinges on the analysis of Iwaidja. Given our results, there is sufficient reason to call into question a reconstruction of a velar contrast in pre-Iwaidja and Proto-Iwaidjan.

Moreover, there is no Iwaidjan-external evidence that supports the reconstruction of an opposition between Proto-Iwaidjan *k and *tʃ. Comparison with forms in neighboring languages shows that Iwaidja has been affected by lenition. This is illustrated in A1, with data from Bininj Gunwok as comparison.

(A1)	IWAIDJA	BININJ GUNWOK
'earthquake'	wula	bula
'spear type'	wirdurrk	birturrk
'spear type'	yalawan	jalaban
'yellow ochre'	karlwa	karlba
'masculine-subsection term'	na-wuyuk	na-kojok
'feminine-subsection term'	ngal-wuyuk	ngal-kojok

As illustrated in A1, the labial and palatal stops lenite in both initial and medial positions. The retroflex stop also lenites in medial position.¹¹ The velar stop lenites to /w/ when there is a following round vowel. Alternations between /k/ and /w/, presumably arising in lenition, are common in Australian languages (Baker 2014:170–75). However, as the 'yellow ochre' correspondence in A1 shows, the velar stop has not otherwise undergone lenition in initial position. Correspondences in medial position, when not followed by a round vowel, are not consistent. As shown in A2, intercontinuant /k/ in Bininj Gunwok sometimes appears as <k> in Iwaidja and sometimes as <h>.

(A2)	IWAIDJA	BININJ GUNWOK
h – k 'boomerang'	wirr h ala	birr k ala
k – k 'matry term'	-wurr k arr	-wurr k arr

However, if there is no contrast between <h> and <k> in Iwaidja, then there is no inconsistency. Both sets in A2 illustrate a /k/ – /k/ correspondence. Overall, therefore, none of the evidence, whether Iwaidja-internal, from congeners within the Iwaidjan family, or from external comparisons, supports the reconstruction of a contrast between *k and *tʃ for Proto-Iwaidjan or a pre-stage of Iwaidja.

The diachronic loss lenition patterns in A1 and A2 have produced complex synchronic patterns of root-initial lenition in paradigms where roots take prefixes: that is, prefixing roots. These complex patterns are conditioned by both phonological and morphological factors (Evans 1997, 2000, 2009). Table A1 provides some examples of prefixing verb paradigms; with transitive verbs, the prefixes cross-reference subject + 3sg object. In these lenition relationships, variants involving complete articulatory closure are fortis; variants not involving closure are lenis—the fortis variants for the palatal, labial, and apicals are <ɟ, b, ld>, and the lenis variants are <y, w, r>.

In the intransitive 'work' paradigm, the lenis allomorph <yama> /jama/ appears after vowels, and the fortis allomorph <jama> /cama/ appears elsewhere, which is consistent with the environment for velar lenition. However, lenition in the transitive paradigms involves an interaction with morphology (columns B–D). For transitive verbs, forms with a third-person subject take one allomorph, and forms with a non-third-person

¹¹ There are no examples of forms in a non-Iwaidjan language with an initial retroflex stop that have a corresponding form in Iwaidja.

	INTRANSITIVE		TRANSITIVE	
	A.	B.	C.	D.
	'work-PRS'	'look for-PRS' (3SG.O)	'hit-PRS' (3SG.O)	'eat-PRS' (3SG.O)
1SG	nga-yama-ng	a-jalma-n	a-bu-n	a-ra
1NSG.EXCL	ngad-jama-ng	ngarra-jalma-n	ngarra-bu-n	ngarru-ra
1NSG.INCL	ad-jama-ng	arra-jalma-n	arra-bu-n	arru-ra
2SG	ang-jama-ng	ku-jalma-n	ka-bu-n	ka-ra
2PL	kud-jama-ng	kurru-jalma-n	kurru-bu-n	kurru-ra
3SG.M	Ø-jama-ng	ri-yalma-n	ri-wu-n	ri-lda
3SG.F	Ø-jama-ng	ka-yalma-n	ka-wu-n	ka-lda
3PL	a-yama-ng	bu-yalma-n	bu-wu-n	bu-lda

TABLE A1. Root-initial lenition.

subject take the other allomorph. Crucially, the distribution of fortis and lenis allomorphs with palatal- and labial-initial roots, such as 'look for' (column B) and 'hit' (column C), is the mirror image of their distribution with apical-initial roots, such as 'eat' (column D). In the palatal 'look for' and labial 'hit' paradigms, the fortis variants <ɟ> and appear with non-third-person subjects, and the lenis variants <y> and <w> appear with third-person subjects. By contrast, in the apical 'eat' paradigm, the fortis variant <ld> appears with third-person subjects, and the lenis variant <r> appears with non-third-person subjects. These lenition patterns belie characterization as continuity lenition because they both are not contrast-preserving and do not consistently enhance the intensity profile of a particular prosodic domain. Crucially, the type of lenis-fortis alternations observed in Table A1 do not occur for velars. Synchronically in Iwaidja, there are no prefixing paradigms with initial velars.

REFERENCES

- BAKER, BRETT. 2014. Word structure in Australian languages. *The languages and linguistics of Australia: A comprehensive guide*, ed. by Harold Koch and Rachel Nordlinger, 139–213. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110279771.139.
- BAUER, LAURIE. 1988. What is lenition? *Journal of Linguistics* 24.381–92. DOI: 10.1017/S002222670001183X.
- BAUER, LAURIE. 2008. Lenition revisited. *Journal of Linguistics* 44.605–24. DOI: 10.1017/S0022226708005331.
- BECKER, MICHAEL; ANDREW NEVINS; and JONATHAN LEVINE. 2012. Asymmetries in generalizing alternations to and from initial syllables. *Language* 88.231–68. DOI: 10.1353/lan.2012.0049.
- BECKMAN, JILL N. 1998. *Positional faithfulness*. Amherst: University of Massachusetts Amherst dissertation.
- BENUS, STEFAN, and ADAMANTIOS I. GAFOS. 2007. Articulatory characteristics of Hungarian 'transparent' vowels. *Journal of Phonetics* 35.271–300. DOI: 10.1016/j.wocn.2006.11.002.
- BOERSMA, PAUL, and DAVID WEENINK. 2013. Praat: Doing phonetics by computer. Version 5.3.53. Online: <http://praat.org>.
- BRAVER, AARON. 2014. Imperceptible incomplete neutralization: Production, non-identifiability, and non-discriminability in American English flapping. *Lingua* 152.24–44. DOI: 10.1016/j.lingua.2014.09.004.
- CAMPBELL-KIBLER, KATHRYN. 2007. Accent, (ING), and the social logic of listener perceptions. *American Speech* 82.32–64. DOI: 10.1215/00031283-2007-002.
- CAMPBELL-KIBLER, KATHRYN. 2009. The nature of sociolinguistic perception. *Language Variation and Change* 21.135–56. DOI: 10.1017/S0954394509000052.
- CAPELL, ARTHUR. 1962. *Some linguistic types in Australia*. Sydney: University of Sydney.
- CAPELL, ARTHUR, and HEATHER HINCH. 1970. *Maung grammar, texts and vocabulary*. The Hague: Mouton.
- CARIGNAN, CHRISTOPHER. 2014. TRACTUS (Temporally Resolved Articulatory Configuration Tracking of Ultrasound) software suite. Online: <http://christophercarignan.github.io/TRACTUS>.
- CARIGNAN, CHRISTOPHER. 2018. Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels. *The Journal of the Acoustical Society of America* 143.2588–2601. DOI: 10.1121/1.5034760.

- CARROLL, JOHN B., and MARGARET N. WHITE. 1973. Word frequency and age of acquisition as determiners of picture-naming latency. *The Quarterly Journal of Experimental Psychology* 25.85–95. DOI: 10.1080/14640747308400325.
- COHEN PRIVA, URIEL. 2017. Informativity and the actuation of lenition. *Language* 93.569–97. DOI: 10.1353/lan.2017.0037.
- COHEN PRIVA, URIEL, and EMILY GLEASON. 2020. The causal structure of lenition: A case for the causal precedence of durational shortening. *Language* 96.413–48. DOI: 10.1353/lan.2020.0025.
- COHEN PRIVA, URIEL, and FLORIAN JAEGER. 2018. The interdependence of frequency, predictability, and informativity in the segmental domain. *Linguistics Vanguard* 4:20170028. DOI: 10.1515/lingvan-2017-0028.
- CONNELL, BRUCE. 2002. Tone languages and the universality of intrinsic F0: Evidence from Africa. *Journal of Phonetics* 30.101–29. DOI: 10.1006/jpho.2001.0156.
- CUTLER, ANNE, and SALLY BUTTERFIELD. 1992. Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language* 31.218–36. DOI: 10.1016/0749-596X(92)90012-M.
- CUTLER, ANNE, and CHARLES CLIFTON, JR. 1984. The use of prosodic information in word recognition. *Attention and performance X: Control of language processes*, ed. by Herman Bouma and Don G. Bowhuis, 183–96. London: Lawrence Erlbaum.
- DAVIDSON, LISA. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America* 120.407–15. DOI: 10.1121/1.2205133.
- DERRICK, DONALD; CATHERINE T. BEST; and ROMAIN FIASSON. 2015. Non-metallic ultrasound probe holder for co-collection and co-registration with EMA. Paper presented at the 18th International Congress of Phonetic Sciences (ICHPS), Glasgow.
- DILLEY, LAURA C., and J. DEVIN MCAULEY. 2008. Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language* 59.294–311. DOI: 10.1016/j.jml.2008.06.006.
- ECKERT, PENELOPE. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12.453–76. DOI: 10.1111/j.1467-9841.2008.00374.x.
- EISNER, FRANK, and JAMES M. MCQUEEN. 2005. The specificity of perceptual learning in speech processing. *Perception & Psychophysics* 67.224–38. DOI: 10.3758/BF03206487.
- ENNEVER, THOMAS; FELICITY MEAKINS; and ERICH R. ROUND. 2017. A replicable acoustic measure of lenition and the nature of variability in Gurindji stops. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8:20. DOI: 10.5334/labphon.18.
- EVANS, NICHOLAS. 1997. Macassan loans and linguistic stratigraphy among the Iwaidjan languages. *Archaeology and linguistics: Global perspectives on ancient Australia*, ed. by Patrick McConvell and Nicholas Evans, 237–59. Melbourne: Oxford University Press.
- EVANS, NICHOLAS. 2000. Iwaidjan, a very un-Australian language family. *Linguistic Typology* 4.91–142. DOI: 10.1515/lity.2000.4.1.91.
- EVANS, NICHOLAS. 2009. Doubled up all over again: Borrowing, sound change and reduplication in Iwaidja. *Morphology* 19.159–76. DOI: 10.1007/s11525-009-9139-4.
- FISCHER, JOHN L. 1958. Social influences on the choice of a linguistic variant. *Word* 14.47–56. DOI: 10.1080/00437956.1958.11659655.
- FOLEY, JAMES. 1970. Phonological distinctive features. *Folia Linguistica* 4.87–92. DOI: 10.1515/flin.1970.4.1-2.87.
- FRALEY, CHRIS, and ADRIAN E. RAFTERY. 2007. Model-based methods of classification: Using the mclust software in chemometrics. *Journal of Statistical Software* 18.1–13. DOI: 10.18637/jss.v018.i06.
- GAHL, SUSANNE. 2008. *Time and thyme* are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84.474–96. DOI: 10.1353/lan.0.0035.
- GICK, BRYAN; DOUGLAS PULLEYBLANK; FIONA CAMPBELL; and NGESSIMO MUTAKA. 2006. Low vowels and transparency in Kinande vowel harmony. *Phonology* 23.1–20. DOI: 10.1017/S0952675706000741.

- GORDON, MATTHEW. 1999. The 'neutral' vowels of Finnish: How neutral are they? *Linguistica Uralica* 35.17–21.
- HALL, KATHLEEN CURRIE. 2009. *A probabilistic model of phonological relationships from contrast to allophony*. Columbus: The Ohio State University dissertation.
- HALLE, MORRIS, and G. N. CLEMENTS. 1983. *Problem book in phonology: A workbook for introductory courses in linguistics and in modern phonology*. Cambridge, MA: MIT Press.
- HAZEN, KIRK. 2008. (ING): A vernacular baseline for English in Appalachia. *American Speech* 83.116–40. DOI: 10.1215/00031283-2008-008.
- HERMES, ANNE; DORIS MÜCKE; and BASTIAN AURIS. 2017. The variability of syllable patterns in Tashlhiyt Berber and Polish. *Journal of Phonetics* 64.127–44. DOI: 10.1016/j.wocn.2017.05.004.
- HEYNE, MATTHIAS; XUAN WANG; KIERAN DORREEN; DONALD DERRICK; and KEVIN WATSON. 2016. The articulation of /ɹ/ in New Zealand English. *The Journal of the Acoustical Society of America* 140.3225. DOI: 10.1121/1.4970185.
- HONEYBONE, PATRICK. 2008. Lenition, weakening and consonantal strength: Tracing concepts through the history of phonology. *Lenition and fortition*, ed. by Joaquim Brandão de Carvalho, Tobias Scheer, and Philippe Ségéral, 9–92. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110211443.1.9.
- HOUSTON, ANN CELESTE. 1985. *Continuity and change in English morphology: The variable (ING)*. Philadelphia: University of Pennsylvania dissertation. Online: <https://repository.upenn.edu/dissertations/AAI8515390>.
- HUEBER, THOMAS; GUIDO AVERSANO; G. CHOLLE; BRUCE DENBY; GÉRARD DREYFUS; YACINE OUSSAR; PIERRE ROUSSEL; and MAUREEN STONE. 2007. Eigentongue feature extraction for an ultrasound-based silent speech interface. Paper presented at the 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing.
- HYMAN, LARRY. 1975. *Phonology: Theory and analysis*. New York: Holt, Rinehart and Winston.
- ISKAROUS, KHALIL; JOYCE McDONOUGH; and D. H. WHALEN. 2012. A gestural account of the velar fricative in Navajo. *Laboratory Phonology* 3.195–210. DOI: 10.1515/lp-2012-0011.
- KATZ, JONAH. 2016. Lenition, perception and neutralisation. *Phonology* 33.43–85. DOI: 10.1017/S0952675716000038.
- KAYE, JONATHAN, and JOHN HARRIS. 1990. Segmental complexity and phonological government. *Phonology* 7.255–300. DOI: 10.1017/S0952675700001202.
- KINGSTON, JOHN. 2008. Lenition. Paper presented to the 3rd Conference on Laboratory Approaches to Spanish Phonology.
- KIRCHNER, ROBERT. 1998. *An effort-based approach to consonant lenition*. Los Angeles: University of California, Los Angeles dissertation. DOI: 10.7282/T3CZ360S.
- KIRCHNER, ROBERT. 2001. Phonological contrast and articulatory effort. *Segmental phonology in optimality theory: Constraints and representations*, ed. by Linda Lombardi, 79–117. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511570582.004.
- KRALJIC, TANYA, and ARTHUR G. SAMUEL. 2006. Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review* 13.262–68. DOI: 10.3758/BF03193841.
- LAVOIE, LISA MARIE. 2001. *Consonant strength: Phonological patterns and phonetic manifestations*. New York: Garland.
- LI, MIN; CHANDRA KAMBHAMETTU; and MAUREEN STONE. 2005. Automatic contour tracking in ultrasound images. *Clinical Linguistics & Phonetics* 19.545–54. DOI: 10.1080/02699200500113616.
- LOHMANN, ARNE. 2018. *Time and thyme* are NOT homophones: A closer look at Gahl's work on the lemma-frequency effect, including a reanalysis. *Language* 94.e180–e190. DOI: 10.1353/lan.2018.0032.
- MAILHAMMER, ROBERT, and MARK HARVEY. 2018. A reconstruction of the Proto-Iwaidjan phoneme system. *Australian Journal of Linguistics* 38.329–59. DOI: 10.1080/07268602.2018.1470455.
- MAILHAMMER, ROBERT; STACEY SHERWOOD; and HYWEL STOAKES. 2020. The inconspicuous substratum: Indigenous Australian languages and the phonetics of stop contrasts in

- English on Croker Island. *English World-Wide* 41.162–92. DOI: 10.1075/eww.00045.mai.
- MAYE, JESSICA; RICHARD N. ASLIN; and MICHAEL K. TANENHAUS. 2008. The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science* 32.543–62. DOI: 10.1080/03640210802035357.
- MCMURRAY, BOB; RICHARD N. ASLIN; and JOSEPH C. TOSCANO. 2009. Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science* 12.369–78. DOI: 10.1111/j.1467-7687.2009.00822.x.
- MIELKE, JEFF; CHRISTOPHER CARIGNAN; and ERIK R. THOMAS. 2017. The articulatory dynamics of pre-velar and pre-nasal /æ/-raising in English: An ultrasound study. *The Journal of the Acoustical Society of America* 142.332–49. DOI: 10.1121/1.4991348.
- MURRAY, ROBERT W. 1988. *Phonological strength and early Germanic syllable structure*. Munich: Wilhelm Fink.
- NIVENS, RICHARD. 1992. A lexical phonology of West Tarangan. *Phonological studies in four languages of Maluku*, ed. by Donald A. Burquest and Wyn D. Laidig, 127–227. Dallas: Summer Institute of Linguistics.
- NOIRAY, AUDE; KHALIL ISKAROUS; and DOUGLAS H. WHALEN. 2014. Variability in English vowels is comparable in articulation and acoustics. *Laboratory Phonology* 5.271–88. DOI: 10.1515/lp-2014-0010.
- NORRIS, DENNIS; JAMES M. MCQUEEN; and ANNE CUTLER. 2003. Perceptual learning in speech. *Cognitive Psychology* 47.204–38. DOI: 10.1016/S0010-0285(03)00006-9.
- OFTEDAL, MAGNE. 1985. *Lenition in Celtic and in Insular Spanish: The secondary voicing of stops in Gran Canaria*. Oxford: Oxford University Press.
- PARKER, STEPHEN. 2008. Sound level protrusions as physical correlates of sonority. *Journal of Phonetics* 36.55–90. DOI: 10.1016/j.wocn.2007.09.003.
- PARKER, STEVE. 2017. Sounding out sonority. *Language and Linguistics Compass* 11: e12248. DOI: 10.1111/lnc3.12248.
- PERKELL, JOSEPH S.; MELANIE L. MATTHIES; MARIO A. SVIRSKY; and MICHAEL I. JORDAN. 1993. Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot ‘motor equivalence’ study. *The Journal of the Acoustical Society of America* 93.2948–61. DOI: 10.1121/1.405814.
- PIERREHUMBERT, JANET B. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46.115–54. DOI: 10.1177/00238309030460020501.
- POPPE, NIKOLAJ NIKOLAEVIČ. 1963. *Tatar manual*. Bloomington: Indiana University Press.
- PRINCE, ALAN, and PAUL SMOLENSKY. 2004. *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.
- PYM, NOREEN, and BONNIE LARRIMORE. 1979. *Papers on Iwaidja phonology and grammar*. Darwin: SIL-AAB.
- RECASENS, DANIEL, and CLARA RODRÍGUEZ. 2016. A study on coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound. *Journal of Phonetics* 59.58–75. DOI: 10.1016/j.wocn.2016.09.002.
- REMIJSEN, BERT. 2013. Tonal alignment is contrastive in falling contours in Dinka. *Language* 89.297–327. DOI: 10.1353/lan.2013.0023.
- ROMERO GALLEGÓ, JOAQUIN. 1996. *Gestural organization in Spanish: An experimental study of spirantization and aspiration*. Storrs: University of Connecticut dissertation. Online: <https://opencommons.uconn.edu/dissertations/AAI9541611>.
- SALTZMAN, ELLIOT L., and KEVIN G. MUNHALL. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1.333–82. DOI: 10.1207/s15326969eco0104_2.
- SAMUEL, ARTHUR G., and TANYA KRALJIC. 2009. Perceptual learning for speech. *Attention, Perception, & Psychophysics* 71.1207–18. DOI: 10.3758/APP.71.6.1207.
- SEZER, ENGİN. 1986. An autosegmental analysis of compensatory lengthening in Turkish. *Studies in compensatory lengthening*, ed. by Leo Wetzels and Engin Sezer, 227–50. Berlin: De Gruyter. DOI: 10.1515/9783110821666-011.
- SHAW, JASON A., and RAHUL BALUSU. 2010. Language contact and phonological contrast: The case of coronal affricates in Japanese loans. *Language contact: New perspectives*, ed. by Muriel Norde, Bob de Jonge, and Cornelius Hasselblatt, 155–80. Amsterdam: John Benjamins.
- SHAW, JASON A.; CATHERINE T. BEST; GERRY DOCHERTY; BRONWEN G. EVANS; PAUL FOULKES; JENNIFER HAY; and KAREN E. MULAK. 2018. Resilience of English vowel

- perception across regional accent variation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 9:11. DOI: 10.5334/labphon.87.
- SHAW, JASON A., and ADAMANTIOS I. GAFOS. 2015. Stochastic time models of syllable structure. *PLoS ONE* 10:e0124714. DOI: 10.1371/journal.pone.0124714.
- SHERWOOD, STACEY. 2015. Indicating and perceiving social hierarchy through language variation: The case of ranuki in Japanese. Sydney: Western Sydney University B.A. honors thesis.
- SINGER, RUTH. 2006. *Agreement in Mawng: Productive and lexicalised uses of agreement in an Australian language*. Melbourne: University of Melbourne dissertation. Online: <http://hdl.handle.net/11343/39232>.
- SKOUSEN, ROYAL. 1975. *Substantive evidence in phonology: The evidence from Finnish and French*. The Hague: Mouton.
- SMITH, CAITLYN. 2018. *Harmony in gestural phonology*. Los Angeles: University of Southern California dissertation.
- SMITH, CAROLINE. 1995. Prosodic patterns in the coordination of consonant and vowel gestures. *Phonology and phonetic evidence: Papers in laboratory phonology 4*, ed. by Bruce Connell and Amalia Arvaniti, 205–22. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511554315.015.
- SODERBERG, CRAIG D., and KENNETH S. OLSON. 2008. Indonesian. *Journal of the International Phonetic Association* 38.209–13. DOI: 10.1017/S0025100308003320.
- SUMMERFIELD, QUENTIN, and MARK HAGGARD. 1977. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *The Journal of the Acoustical Society of America* 62.435–48. DOI: 10.1121/1.381544.
- SUMNER, MEGHAN, and ARTHUR G. SAMUEL. 2009. The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language* 60.487–501. DOI: 10.1016/j.jml.2009.01.001.
- ÜNAL-LOGACEV, ÖZLEM; MARZENA ŻYGIS; and SUSANNE FUCHS. 2017. Phonetics and phonology of soft ‘g’ in Turkish. *Journal of the International Phonetic Association* 49.183–206. DOI: 10.1017/S0025100317000317.
- VENNEMANN, THEO. 1988. *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin*. Berlin: Mouton de Gruyter.
- WARNER, NATASHA; ALLARD JONGMAN; JOAN SERENO; and RACHÈL KEMPS. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32.251–76. DOI: 10.1016/S0095-4470(03)00032-9.
- WHALEN, DOUGLAS H.; KHALIL ISKAROUS; MARK K. TIEDE; DAVID J. OSTRY; HEIKE LEHNERT-LEHOULLIER; ERIC VATIKIOTIS-BATESON; and DONALD S. HAILEY. 2005. The Haskins optically corrected ultrasound system (HOCUS). *Journal of Speech, Language, and Hearing Research* 48.543–53. DOI: 10.1044/1092-4388(2005/037).
- WHALEN, DOUGLAS H., and ANDREA G. LEVITT. 1995. The universality of intrinsic F₀ of vowels. *Journal of Phonetics* 23.349–66. DOI: 10.1016/S0095-4470(95)80165-0.
- YIGEZU, MOGES, and GERRIT J. DIMMENDAAL. 1998. Notes on Baale. *Surmic languages and cultures*, ed. by Gerrit J. Dimmendaal and Marco Last, 273–317. Cologne: Rüdiger Köppe.
- YU, YONGJIAN, and SCOTT T. ACTON. 2002. Speckle reducing anisotropic diffusion. *IEEE Transactions on Image Processing* 11.1260–70. DOI: 10.1109/TIP.2002.804276.

Shaw
Yale University, Department of Linguistics
370 Temple Street
New Haven, CT 06511
[jason.shaw@yale.edu]
[c.carignan@ucl.ac.uk] (Carignan)
[tga932@uowmail.edu.au] (Agostini)
[R.Mailhammer@westernsydney.edu.au] (Mailhammer)
[mark.harvey@newcastle.edu.au] (Harvey)
[donald.derrick@canterbury.ac.nz] (Derrick)

[Received 7 November 2018;
revision invited 27 January 2019;
revision received 22 July 2019;
revision invited 31 October 2019;
revision received 30 January 2020;
accepted with revisions 17 February 2020;
revision received 27 March 2020;
accepted 27 March 2020]